MANTEL–HAENSZEL–TYPE INFERENCE FOR ODDS RATIOS WITH ORDINAL RESPONSES

By

I-MING LIU

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1995

© Copyright 1995

by

I-Ming Liu

To my parents

and

Hwarng-Du Hwang

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor Alan Agresti for his wisdom while guiding me towards my degree. His constant encouragement provided me with the energy to carry on. Moreover, his work ethic showed me how to enjoy and perform research. Special thanks go to Professor Ramon Littell for serving on my committee and for the valuable experience I gained from him during the two years I worked in IFAS. I would like to express my appreciation for Professor Mark Yang, who kindly offered me my first research assistantship at the University of Florida. A note of thanks goes to Professor Andrew Rosalsky for removing the obstacles to my admission to this department in a very short period of time. I would also like to thank Professors Malay Ghosh and Myron Chang of the Department of Statistics and Professor James Algina of the Department of Foundations of Education, as well as the faculty members in the department for imparting their vast knowledge to me when I was a graduate student here.

I would like to thank my family for all of the support they have given me during my study. Thanks also go to my fiancé, Hwarng-Du Hwang, who has been patiently waiting in Taiwan for the completion of my dissertation. Finally, to my friends, I wish to acknowledge my pleasure of knowing them.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv

ABSTRACT vii

CHAPTERS

1 INTRODUCTION 1

1.1 A Brief Introduction to the Problem 1
1.2 Outline of the Ordinary Mantel–Haenszel Method 3
1.3 Outline of the Generalized Mantel–Haenszel Method 6
1.4 Format of Dissertation 8

2 ORDINAL ODDS RATIO ESTIMATION 11

2.1 Introduction 11
2.2 An Ordinal Odds Ratio Estimator 12
2.3 Asymptotics Under the Sparse–Data Limiting Model 19
2.4 Asymptotics Under the Large–Stratum Limiting Model 33
2.5 Ordinal Test of Conditional Independence 42
2.6 An Example 46

3 EVALUATION OF ORDINAL ODDS RATIO ESTIMATORS 50

3.1 Introduction 50
3.2 Efficiency 52
3.3 Simulation Study 71
3.4 Tests of the Common Cumulative Odds Ratio Assumption 82
3.5 Weighted Average of the Stratum–Specific Cumulative Odds Ratio 89

4 EXTENSIONS OF THE ORDINAL ODDS RATIO ESTIMATION . 91

4.1 Introduction 91
4.2 Ordinal–Nominal Cumulative Odds Ratios for $r \times c \times K$ Contingency Tables 93

4.3 Ordinal–Ordinal Cumulative Odds Ratios for $r \times c \times K$ Contingency Tables 116
4.4 Alternative Ordinal Odds Ratios for $2 \times c \times K$ Contingency Tables 121

5 HETEROGENEOUS ORDINAL ODDS RATIOS ESTIMATION 128

5.1 Introduction 128
5.2 Heterogeneous Odds Ratios for a Binary Response 130
5.3 Heterogeneous Odds Ratios for an Ordinal Response 132

6 CONCLUSION 146

6.1 Summary of Results 146
6.2 Future Research 149

APPENDICES

A CALCULATIONS FOR SECTION 2.4 152

B CALCULATIONS FOR SECTION 2.6 162

REFERENCES 166

BIOGRAPHICAL SKETCH 170

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

MANTEL–HAENSZEL–TYPE INFERENCE FOR ODDS RATIOS
WITH ORDINAL RESPONSES

By

I-Ming Liu

May 1995

Chairman: Alan G. Agresti
Major department: Statistics

The Mantel–Haenszel estimator for summarizing association from several 2×2 contingency tables is simple, yet has good asymptotic properties for two limiting situations – (1) when the sample size within each stratum becomes large and the number of strata is fixed, and (2) when the number of strata increases with the sample size. This dissertation discusses estimating odds ratios for an ordinal response. For a proportional odds model for $2\times c \times K$ contingency tables in which the columns are ordered, we provide a simple extension of the Mantel-Haenszel estimator. Our estimator, unlike the maximum likelihood (ML) estimator, is consistent under both limiting models. We also obtain an estimated standard error for our estimator that is consistent under both limiting situations. Results of a simulation study reveal that, for sparse data, the expected value of the ML estimator for the log ordinal odds ratio exceeds the true parameter, but the expected value of the proposed estimator is very close to the true parameter. Also, the proposed estimator does not usually have much loss of efficiency compared to the ML estimator under the limiting situation (1).

Some related problems are discussed, such as a test of the common ordinal odds ratio assumption and an ordinal test of conditional independence. Extensions of results are also given to $r \times c \times K$ contingency tables for either r nominal rows or r ordinal rows. Furthermore, two types of generalized linear mixed model with an ordinal response are introduced.

CHAPTER 1
INTRODUCTION

1.1 A Brief Introduction to the Problem

The Mantel–Haenszel (1959) estimator is a popular way of summarizing the association from several 2×2 contingency tables, particularly when the data are sparse. It is commonly used when we assume a common odds ratio in those tables. It is also sometimes used when the association is not expected to vary drastically among the tables, and we want a summary of the associations from the various tables.

For $2\times2\times K$ contingency tables, we can obtain an alternative estimator based on the assumption of a constant odds ratio for the K strata of 2×2 tables by fitting a logit model or the equivalent loglinear model. We treat one of the binary classifications (the columns) as the response, the other one (the rows) and the K strata as factors, and fit a logit model with main effects but no interaction. The corresponding loglinear model is the model of no three-factor interaction. When the model holds but K is large and the data are sparse, the maximum likelihood (ML) estimator based on the model tends to overestimate the true odds ratio. An important application of this type is when the number of tables grows with the sample size, such as when each table presents results for a matched pair. Standard asymptotic properties of unconditional maximum likelihood estimators fail when the number of parameters grows at the same rate as the sample size (Neyman and Scott, 1948). When each stratum consists of a single matched pair, for instance, the unconditional ML estimator of the common

odds ratio converges to the square of the true common odds ratio (Andersen, 1973, p. 69; Andersen, 1980, p. 244).

The Mantel–Haenszel estimator is simple, yet has good asymptotic properties for the two standard types of asymptotics – (1) when the sample size within each stratum becomes large and the number of strata is fixed and (2) when the number of strata increases with the sample size. We call limiting situation (1) a *large–stratum* limiting model, or *Model I*. We call limiting situation (2) a *sparse–data* limiting model, or *Model II*.

The Mantel–Haenszel estimator discussed is for a binary response case. Here, we are interested in estimation of odds ratios that occur naturally when the response is ordinal, rather than binary. Our methods will apply to $2 \times c \times K$ contingency tables in which the c columns are ordered. Such data are quite common. For instance, we might want to compare response distributions for two treatments using an ordered categorical response such as (poor, fair, good, excellent), when data are available from several centers.

One way to analyze the ordinal response data is by fitting a proportional odds model (McCullagh, 1980). A common ordinal odds ratio can be obtained from a proportional odds model with no interaction between the row and the control factors. As in the case of the logit model with a binary response, the ML estimator of the parameter in the proportional odds model may be poor when the data are sparse. We provide a simple extension of the Mantel-Haenszel estimator. We show that that estimator, unlike the ML estimator, is consistent under both limiting models. In addition, we extend the Robins *et al.* (1986) method to obtain an estimated standard error for our estimator that is consistent under both limiting situations. We will also discuss some related problems, such as tests of conditional independence, tests of the common odds ratio assumption, efficiency and simulation studies for the

Mantel–Haenszel–type estimator compared to the ML estimator, and extensions of the proposed estimator when the row factor has r nominal or ordinal categories.

1.2 Outline of the Ordinary Mantel–Haenszel Method

For a binary response case, we call the Mantel–Haenszel estimator the *ordinary* Mantel–Haenszel estimator. We begin by reviewing some existing methodologies related to the ordinary Mantel–Haenszel method, including a common odds ratio estimate and tests of conditional independence. Also, Kuritz *et al.* (1988) reviewed the Mantel–Haenszel test and estimator and their generalizations.

For $2 \times 2 \times K$ contingency tables, a common odds ratio estimator can be obtained by fitting an additive logit model with main effects but no interaction using the ML method. Alternatively, Woolf (1955) estimated the common odds ratio using the exponent of a weighted average of sample log odds ratios from the separate strata. However, compared to these two methods, the ordinary Mantel–Haenszel estimator is less affected by strata with small sample sizes. The estimator is consistent under both Model I (Gart, 1962) and Model II (Breslow, 1981).

The ML estimator for the logit model under Model II is a poor estimator because the number of nuisance parameters corresponding to the K strata grows at the same rate as the sample size. In this case, we can eliminate the nuisance parameters by conditioning on margins for each 2×2 contingency table. The conditional distribution is noncentral hypergeometric in each stratum. The ML estimator for the common odds ratio based on the conditional distribution is also consistent under both limiting cases. Liang (1985) pointed out a difference, however, between properties of the ordinary Mantel–Haenszel estimator and the conditional ML estimator. He showed that when subject responses are correlated within tables, the Mantel–Haenszel estimator is consistent, but the conditional ML estimator is not consistent.

There are several estimators for the variance of the ordinary Mantel–Haenszel estimator. Hauck (1979) proposed a variance under the large–stratum limiting model. Breslow (1981) proposed another variance under the sparse–data limiting model. Robins *et al.* (1986) proposed an estimator for the variance having broader scope, being consistent under either Model I or Model II. An estimator that is consistent under either Model I or Model II is called a *dually* consistent estimator. In this dissertation, we construct dually consistent estimators.

Mantel and Haenszel (1959) also gave a test based on the hypergeometric distribution comparing two groups on a binary response, adjusting for control variables. Under conditional independence, the test statistic is approximately chi–squared distributed under both limiting models. The sample size discussions for the test are available in Mantel and Fleiss (1980), Breslow and Day (1980, p. 139), Fleiss (1981), Koch *et al.* (1985), and Agresti (1990, p. 232). This test is inappropriate when the association changes dramatically across the strata. There is a relationship between the ordinary Mantel–Haenszel estimate and the test statistic. Mantel and Haenszel (1959) showed that the estimate equals unity only when the estimated test statistic equals zero. So, the test statistic was conducted as a significance test of the departure of the weighted average of the stratum–specific odds ratio from unity.

The assumption for the ordinary Mantel–Haenszel estimator is of a common odds ratio; that is, the odds ratio is constant across the strata. When this assumption fails, the Mantel–Haenszel estimate is a summary of the odds ratios among the strata. Without the assumption, we can use the Mantel–Haenszel estimate only under the large–strata situation and with the estimated standard error proposed by Guilbaud (1983), since the dually consistent variance estimator of Robins *et al.* (1986) fails without the assumption. Next, we will review several methods of testing the homogeneous odds ratio assumption.

One possible hypothesis test can be conducted by testing goodness–of–fit of the additive logit model using the ML method, provided the within–stratum sample sizes are adequate. The test is equivalent to a test of no interaction between the row and the control factors. The goodness–of–fit test statistic has an asymptotically chi–squared distribution with $K-1$ degrees of freedom (df), when the additive logit model holds.

Another test was provided by Breslow and Day (1980). They developed a test for the homogeneous odds ratio assumption that does not require fitting logit models. The Breslow–Day test statistic is focused directly on the potential lack of homogeneity of odds ratios. It sums up the squared deviations of observed and fitted values, each standardized by its variance. All fitted values and their variances are calculated based on treating the ordinary Mantel–Haenszel estimate as the true common odds ratio. They claimed that the test statistic follows an approximate chi–squared distribution with df= $K-1$ if the homogeneity assumption is valid and the sample size within each stratum is large. Tarone (1985) found that the Breslow–Day test statistic is not asymptotically chi–squared. He proved that the Breslow–Day test statistic is stochastically larger under the homogeneity assumption than a chi–squared random variable with df= $K-1$, and he developed a correction method so that the new statistic follows asymptotically a chi–squared distribution.

Both the Breslow–Day type test and the test of goodness–of–fit for the additive logit model require large sample sizes within each stratum. These tests are not appropriate under Model II. When the number of strata increases with the sample size, Liang and Self (1985) proposed a score test for homogeneous odds ratios under the assumption that the log odds ratios across the strata are independent and identically distributed with a distribution. They used the efficient score statistic from the conditional likelihood to conduct a test of the hypothesis.

Also, Liu and Pierce (1993) assumed that the log odds ratios across the strata are a sample from a population with unknown mean and variance. They were interested

in obtaining inferences regarding the mean and the variance by direct computation of the conditional likelihood function for the mean and the variance. Then, a test of the common odds ratio assumption can be established by testing that the variance of the log odds ratios equals zero. Their method is more general than the Liang and Self (1985) method, since it describes the heterogeneity of the log odds ratios across the strata.

1.3 Outline of the Generalized Mantel–Haenszel Method

In Section 1.2, we discussed the Mantel–Haenszel method for a binary response case. In this section, we will review some generalizations of this method for a multiple response case. Mantel–Haenszel type estimates for odds ratios have been proposed when the response has c nominal categories. Also, the Mantel–Haenszel test of conditional independence was generalized to $2 \times c \times K$ contingency tables and to $r \times c \times K$ contingency tables for either c nominal or ordinal response categories.

First, we review the generalized Mantel–Haenszel estimator for $2 \times c \times K$ contingency tables having c nominal response categories. One possible Mantel–Haenszel partial estimator for particular columns j and h can be obtained using the ordinary Mantel–Haenszel method by considering those two columns only. Mickey and Elashoff (1985) proposed a more efficient generalized Mantel–Haenszel partial estimator using information from all pairs of columns. Furthermore, Greenland (1989) provided dually consistent variance and covariance estimators for the Mickey and Elashoff estimators. Liang (1987) extended the Mickey and Elashoff method by introducing a class of estimating functions. Sato (1991) proposed dually consistent variance and covariance estimators for the partial odds ratios estimators produced from Liang's (1987) estimating functions.

Recently, Yanagawa and Fujii (1994) introduced a projection method to generalize the Mantel–Haenszel estimator to $2 \times c \times K$ contingency tables having c nominal

response categories. The method projects the space of the log Mantel–Haenszel estimators obtained from all K 2×2 subtables onto a linear space of the log odds ratios. The Mickey–Elashoff and Greenland estimators are special cases of their estimators. Also, their estimators are asymptotically equivalent to Liang's extended Mantel–Haenszel estimators by weighting appropriately. The conditional ML estimators for this case can be obtained by conditioning on the marginal totals for each $2 \times c$ contingency table. The counts follow conditionally the noncentral multiple hypergeometric distribution (Plackett 1981, p. 48). As in the binary response case, the conditional ML estimators are also dually consistent, but they may have an excessive computational burden (Mickey and Elashoff, 1985). There is no software available for obtaining the conditional ML estimates in this case.

For a test of conditional independence, Birch (1965), Landis *et al.* (1978), and Mantel and Byar (1978) generalized the Mantel–Haenszel test statistic to $r \times c \times K$ contingency tables. Conditional on margins for each $r \times c$ contingency table, the counts follow a multiple hypergeometric distribution under the assumption of conditional independence. When the row factor and the response are nominal, the test statistic is a quadratic form having an asymptotic chi–squared distribution with df= $(r-1)(c-1)$ under conditional independence for both Models I and II. When the response and the row factor are ordinal, we can assign scores to the response categories and to the row factor categories. The test statistic, which is asymptotically chi–squared with df=1, was provided by Mantel (1963).

Yanagawa and Fujii (1990) generalized the Breslow–Day (1980) test for the common odds ratio assumption to test the homogeneity of partial odds ratios for $2 \times c \times K$ contingency tables. Following Tarone's (1985) correction term for the Breslow–Day test statistic for the binary response case, Yanagawa and Fujii used adjusted generalized Mantel–Haenszel partial estimators to create a Breslow–Day type test having

asymptotically a chi–squared distribution with df= $(K-1)(c-1)$ under the null hypothesis.

For an ordinal response, surprisingly the Mantel–Haenszel estimator has not been generalized. A common ordinal odds ratio for $2 \times c \times K$ contingency tables can be estimated by fitting a proportional odds model or an adjacent–category logit model assuming no interaction between the row variable and the strata, using the ML method. However, the ML estimators for the ordinal odds ratio in both models may not be good estimators when the data are sparse. Since the proportional odds model is not a canonical link model (*i. e.*, it does not have reduced sufficient statistics), the nuisance parameters cannot be eliminated by conditioning on the marginal totals for each $2 \times c$ contingency table. Therefore, there is no conditional ML estimator for the common ordinal odds ratio based on the proportional odds model.

McCullagh and Nelder (1989, p. 273) proposed a pseudo 'conditional likelihood' estimate for the ordinal odds ratio based on a proportional odds model using an estimating equation, for the one–stratum case. The estimating equation behaves like a log conditional likelihood derivative. It is easy to extend their method to the K–stratum case. They provided a computation method that involves the inverse of the cell means. For this reason, their method cannot be used with extremely sparse data, such as a case–control study in which each stratum consists of a single matched pair of subjects, since the cell means may be zero. In this dissertation, we propose an alternative common ordinal odds ratio estimator using the Mantel–Haenszel method that is dually consistent.

1.4 Format of Dissertation

In Chapter 2, we introduce our ordinal Mantel–Haenszel–type estimator, and the model on which it is based. We also generalize the Robins *et al.* (1986) dually consistent estimator to obtain an estimated variance for the proposed estimator. The

proofs of consistency are given under each limiting model. Then, we present a test of conditional independence using the Mantel–Haenszel method. For the test of conditional independence, we have a similar finding as in the binary response case. The estimated test statistic equals zero when the proposed estimate equals unity. Finally, we apply our generalized Mantel–Haenszel type estimate to a sparse–data example in which each stratum consists of a single matched pair of subjects.

Having shown that the proposed estimator is dually consistent, we study the performance of the estimator compared to the ML estimator in Chapter 3. Under the large–stratum case, the ML estimators have good asymptotic properties (Lehmann, 1983, Chapter 6). They are asymptotically efficient under Model I when a proportional odds model holds. In Section 3.2, we use the Pitman asymptotic relative efficiency (ARE) to describe the loss of efficiency for the proposed estimator compared to the ML estimator under Model I. The results show that our estimator does not suffer much loss of efficiency. Furthermore, we conduct a simulation study to show that the proposed estimator is much better than the ML estimator under the sparse–data case. Section 3.4 provides some tests for the common ordinal odds ratio assumption. Finally, we describe an average ordinal odds ratio estimator when the common ordinal odds ratio assumption fails.

In Chapter 4, we extend the proposed ordinal odds ratio estimator to $r \times c \times K$ contingency tables having either r nominal or ordinal row categories. In addition, we present an alternative Mantel–Haenszel–type ordinal odds ratio estimator based on an adjacent–category logit model.

Since the ordinal odds ratio may not be constant across the strata, it is reasonable to assume that the log ordinal odds ratios across the strata behave like a sample from a population with unknown mean and variance. Then, we can describe the heterogeneity by the variance of the log ordinal odds ratios. Chapter 5 studies heterogeneous ordinal odds ratio estimation based on generalized linear mixed models. In Chapter

6, we summarize the results of this dissertation and we discuss some further research topics for future work.

CHAPTER 2
ORDINAL ODDS RATIO ESTIMATION

2.1 Introduction

The Mantel–Haenszel (1959) estimator is well developed for summarizing the association from several 2×2 contingency tables. It is simple, yet has good asymptotic behavior under the large–stratum limiting model (Model I) and the sparse–data limiting model (Model II). This chapter considers estimation of ordinal odds ratios. Our methods apply to $2 \times c \times K$ contingency tables in which the c columns are ordered. For analyzing the ordinal response data, we can summarize the association from several $2 \times c$ contingency tables assuming a constant ordinal odds ratio across the tables. The ordinal odds ratios are based on a proportional odds model assuming no interaction between the row and the control factors. As in the binary response case, the model–based ML estimator may be poor under Model II. Thus, for the common ordinal odds ratio, we present a simple extension of the Mantel-Haenszel estimator. We show that that estimator, unlike the ML estimator, is consistent under both limiting models. We also extend the Robins *et al.* (1986) method to obtain an estimated standard error for our estimator that is dually consistent.

In addition, we generalize the Mantel–Haenszel (1959) test of conditional independence to the comparison of two groups on an ordinal response, adjusting for the control factor. When the proportional odds model holds, our result shows that we can simply use a test proposed by Mantel (1963) for testing conditional independence.

Section 2.2 introduces our ordinal Mantel-Haenszel–type estimator and the model on which it is based. Section 2.3 shows that the proposed estimator and the proposed

estimated standard error are consistent under Model II. Section 2.4 shows that the proposed estimator and the same proposed estimated standard error are consistent under Model I. Section 2.5 provides a test of conditional independence. Section 2.6 gives an example, in which we compare the unaided distance vision distributions of the left and right eyes, adjusting for each patient. For this example, each stratum consists of a single matched pair.

2.2 An Ordinal Odds Ratio Estimator

We now introduce the proportional odds model for analyzing ordinal response data for $2 \times c \times K$ contingency tables. Let Y denote a response variable having c ordered categories. Let Z denote an explanatory variable having two categories, and let W denote a control variable having K categories. Also, we let π_{ijk} denote the probability that the response Y falls at level $j \in \{1,\ldots,c\}$, when the explanatory variable Z is at level $i \in \{1,2\}$ and the control variable W is at level $k \in \{1,\ldots,K\}$. The probabilities $\{\pi_{ijk}\}$ satisfy $\pi_{i1k} + \cdots + \pi_{ick} = 1$ for all $i = 1,2$ and $k = 1,\ldots,K$. The jth *cumulative* probability is denoted by $\pi^*_{ijk} = \pi_{i1k} + \cdots + \pi_{ijk}$, for all $i = 1,2$, $j = 1,\ldots,c-1$ and $k = 1,\ldots,K$. Then, we consider a proportional odds model having form

$$\log\left(\frac{\pi^*_{ijk}}{1-\pi^*_{ijk}}\right) = \alpha_{jk} + \Delta I_{\{i=1\}}\ ,\ i = 1,2,\ j = 1,\ldots,c-1,\ k = 1,\ldots,K, \quad (2.1)$$

where $I_{\{i=1\}}$ is an indicator variable for which $I_{\{i=1\}}$=1 when $i = 1$ and $I_{\{i=1\}} = 0$ when $i = 2$.

In the proportional odds model (2.1), at each level of W, the odds for level 1 of Z that the response falls below level j are $\exp(\Delta)$ times those for level 2 of Z. Note that $\exp(\Delta)$ is an odds ratio utilizing cumulative probabilities. It is sometimes referred to as a *cumulative odds ratio* (Agresti, 1990, p. 322). The *proportional odds* property refers to the common value of this odds ratio for all "cutpoints" $j = 1,\ldots,c-1$

(Agresti, 1990, p. 322). Since this type of model considers the cumulative probabilities and the logit link, it is also called a *cumulative logit model.* Since model (2.1) has the additional simplification of a common cumulative odds ratio for each stratum of Z, we call it a *common proportional odds model.* In the common proportional odds model (2.1), we refer to $\psi = \exp(\Delta)$ as a *common cumulative odds ratio.*

In the binary response case, Chapter 1 noted that the common odds ratio can be obtained by fitting an additive logit model assuming a constant odds ratio across the K strata. When the response variable Y has two categories, the common proportional odds model (2.1) simplifies to that logit model; that is, the common odds ratio for the binary logit model is a special case of the common cumulative odds ratio for the model (2.1). For a binary response, the odds for level 1 of Z that the response Y falls in level 1 are $\exp(\Delta)$ times the odds for level 2 of Z, at each level of W. When we collapse the response into a binary outcome ($\leq j$, $> j$) for $j \in \{1, \ldots, c-1\}$, the common proportional odds model (2.1) has the same odds ratio for all strata. Therefore, the common proportional odds model (2.1) is an extension that implies a common odds ratio across levels of W for each of the $(c-1)$ ways of collapsing the ordinal response to a binary response.

Now, we introduce notations for $2 \times c \times K$ contingency tables. We assume that each of the $2 \times c$ contingency tables for the K strata is formed by two independent multinomial observations $(X_{1jk}, X_{2jk}, j = 1, \ldots, c)$, with sample sizes denoted by n_{1k} and n_{2k}, for $k = 1, \ldots, K$. The corresponding multinomial probabilities for $\{X_{ijk}, i = 1, 2, j = 1, \ldots, c, k = 1, \ldots, K\}$ are $\{\pi_{ijk}, i = 1, 2, j = 1, \ldots, c, k = 1, \ldots, K\}$. Denote the total sample size within table k by $N_k = n_{1k} + n_{2k}$, and the total sample size over the strata by $N_+ = \sum_k N_k$. Also, denote the column totals by m_{jk}, for $j = 1, \ldots, c$ and $k = 1, \ldots, K$. Table 2.1 illustrates the notations for a stratum k.

Table 2.1. A $2 \times c$ contingency table for the kth stratum

level i for the	level j for the response variable Y				
explanatory variable Z	1	2	...	c	Total
1	X_{11k}	X_{12k}	...	X_{1ck}	n_{1k}
	(π_{11k})	(π_{12k})	...	(π_{1ck})	(1)
2	X_{21k}	X_{22k}	...	X_{2ck}	n_{2k}
	(π_{21k})	(π_{22k})	...	(π_{2ck})	(1)
Total	m_{1k}	m_{2k}	...	m_{ck}	N_k

When $c = 2$, Mantel and Haenszel (1959) proposed a non-model-based estimator of the common odds ratio

$$\hat{\psi}_{\mathrm{MH}} = \frac{\sum_{k=1}^{K}[X_{11k}X_{22k}/N_k]}{\sum_{k=1}^{K}[X_{12k}X_{21k}/N_k]}. \tag{2.2}$$

This is the ordinary Mantel–Haenszel estimator for the common odds ratio. The ordinary Mantel–Haenszel estimator has good asymptotic properties under both the sparse-data and the large-stratum limiting models. We now formally define this sparse-data limiting model (Model II) as the case in which $K \to \infty$ as $N_+ \to \infty$ with $N_+/K = O(1)$, and $\{N_k, k = 1, \ldots, K\}$ are bounded. Also, the large-stratum limiting model (Model I) is the case in which $N_k \to \infty$ but the number of tables K is fixed. Under Model II, the asymptotic properties of the unconditional ML estimator of the common odds ratio obtained by fitting the logit model fail (Neyman and Scott, 1948). The estimator deteriorates as the sample size per partial table decreases. When each stratum consists of a single matched pair, the unconditional ML estimator of the common odds ratio converges to the square of the true common odds ratio (Andersen, 1973, p. 69; Andersen, 1980, p. 244). However, both the ordinary Mantel–Haenszel estimator and the unconditional ML estimator have good

asymptotic properties under the large–stratum limiting model (Model I), and the ML estimator is asymptotically efficient in this setting.

For fixed j, the common proportional odds model (2.1) implies the logit model for a collapsing of the response into a binary outcome $(\leq j, > j)$. Thus, when the model holds, the logit model holds for the collapsed response scale, for any j. However, the unconditional ML estimator of the common odds ratio in the logit model is not consistent when the number of parameters grows at the same rate as the sample size. This implies that when the number of strata is large and the data are sparse, the unconditional ML estimator of the common cumulative odds ratio, $\exp(\hat{\Delta})$ in the common proportional odds model (2.1), may be poor.

We next motivate a simple estimator of the common cumulative odds ratio that, like the Mantel–Haenszel estimator for $2 \times 2 \times K$ contingency tables, behaves well for sparse data with many strata or for large sample sizes within each of a fixed number of strata. If we collapse K $2 \times c$ contingency tables appropriately such that we can apply the ordinary Mantel–Haenszel method for a binary response case, we then have a good estimator under both limiting situations. The common proportional odds model (2.1) shows that the odds ratios are the same for all collapsings of the response into a binary outcome $(\leq j, > j)$ for $j \in \{1, \ldots, c-1\}$ and for all strata. If the common proportional odds model (2.1) holds, the logit models also hold for all collapsings.

Liang and Zeger (1986) introduced a class of estimating equations for repeated measurement data in a longitudinal data analysis. They claimed that under a naive assumption that repeated observations by a subject are independent of one another, their approach still leads to consistent estimators if the regression model is correctly specified. A disadvantage of the estimators is that they may have some loss of efficiency. Using a similar approach, we will naively treat different collapsed tables for each stratum as independent samples to obtain a simple estimator. Later, we will

adjust the variance of the simple estimator by considering covariances between the collapsings.

Under the naive independence assumption, a $2 \times c$ contingency table is viewed as $c-1$ 2×2 contingency tables, with all collapsed tables being independent. We apply the same argument to each of the K strata. So, the $2 \times c \times K$ contingency table can then be viewed as a $2 \times 2 \times K(c-1)$ contingency table. Denote the cumulative count for the jth collapsing of the kth $2 \times c$ table by $X^*_{ijk} = X_{i1k} + \cdots + X_{ijk}$, for $i = 1, 2$, $j = 1, \ldots, c-1$ and $k = 1, \ldots, K$. Let $R_{jk} = X^*_{1jk}(n_{2k} - X^*_{2jk})/N_k$ and $S_{jk} = (n_{1k} - X^*_{1jk})X^*_{2jk}/N_k$. Thus, the proposed Mantel–Haenszel–type estimator for the common cumulative odds ratio is

$$\hat{\psi} = \frac{\sum_{k=1}^{K}\sum_{j=1}^{c-1} X^*_{1jk}(n_{2k} - X^*_{2jk})/N_k}{\sum_{k=1}^{K}\sum_{j=1}^{c-1}(n_{1k} - X^*_{1jk})X^*_{2jk}/N_k} = \frac{\sum_{k=1}^{K}\sum_{j=1}^{c-1} R_{jk}}{\sum_{k=1}^{K}\sum_{j=1}^{c-1} S_{jk}}. \tag{2.3}$$

Much of this dissertation relates to properties of this estimator. Since the estimator based on the naive assumption may have a loss of efficiency, we study efficiency in Chapter 3.

Another motivation of the simple estimator (2.3) can be obtained using an unbiased estimating function. For a 2×2 contingency table, conditional on row and column margins n_1, n_2, m_1, and m_2, the count X_{11} has a noncentral hypergeometric distribution with odds ratio ψ (Breslow and Day 1980, p. 240). The expectation of $X_{11}X_{22}$ equals $\psi \mathrm{E}(X_{12}X_{21})$ (Breslow, 1981). McCullagh and Nelder (1989, p. 355) introduced an unbiased estimating function for ψ over the K independent 2×2 contingency tables as follows:

$$\sum_k \frac{X_{11k}X_{22k} - \psi X_{12k}X_{21k}}{N_k}. \tag{2.4}$$

The estimator produced by the function (2.4) is the ordinary Mantel–Haenszel estimator (2.2). A similar argument of estimating functions based on conditional distributions is also given by Liang (1987).

We now generalize the unbiased estimating function (2.4) to the ordinal response case. Denote the cumulative column total by $m_{jk}^* = m_{1k} + \cdots + m_{jk}$ for $j = 1, \ldots, c-1$ and $k = 1, \ldots, K$. For the jth collapsed $2 \times c$ contingency table, conditional on margins n_{1k}, n_{2k}, m_{jk}^*, and $(N_k - m_{jk}^*)$, X_{1jk}^* has the noncentral hypergeometric distribution with odds ratio ψ for $j \in \{1, \ldots, c-1\}$ and $k \in \{1, \ldots, K\}$. The expectation of $X_{1jk}^*(n_{2k} - X_{2jk}^*)$ equals $\psi \mathrm{E}((n_{1k} - X_{1jk}^*)X_{2jk}^*)$ by the Breslow (1981) argument. Thus, an unbiased estimating function for ψ can be written as

$$\sum_{k=1}^{K} \sum_{j=1}^{c-1} \frac{X_{1jk}^*(n_{2k} - X_{2jk}^*) - \psi(n_{1k} - X_{1jk}^*)X_{2jk}^*}{N_k} . \tag{2.5}$$

The estimator produced by the unbiased estimating function (2.5) is the proposed Mantel–Haenszel–type estimator (2.3) for the common cumulative odds ratio.

For a binary response, the conditional ML estimator for the common odds ratio is an alternative dually consistent estimator to the ordinary Mantel–Haenszel estimator. However, the dually consistent conditional ML estimator does not exist for the common cumulative odds ratio ψ. We first discuss the reason and then introduce a pseudo 'conditional likelihood' estimator proposed by McCullagh and Nelder (1989). For each collapsed table, conditioning on the margins, X_{1jk}^* is distributed as the noncentral hypergeometric distribution. Over all $c-1$ collapsings, the noncentral hypergeometric distributions depends on the parameter ψ only. Unfortunately, there is no joint conditional distribution of $\{X_{1jk}^*\}$ that depends only on ψ because the distribution also depends on the nuisance parameters $\{\alpha_{jk}\}$. Therefore, the nuisance parameters cannot be eliminated by conditioning on the margins. Under the sparse–data limiting model, if there are nuisance parameters, standard asymptotic properties for the ML estimators then fail, because the number of nuisance parameters $\{\alpha_{jk}\}$ grows as $K \to \infty$. On the other hand, there is no reduced sufficient statistics for the nuisance parameters based on the common proportional odds model (2.1). Thus,

the method for eliminating the nuisance parameters by conditioning on the reduced sufficient statistics does not work in this case.

McCullagh and Nelder (1989) proposed an estimating function to obtain a pseudo 'conditional likelihood' estimator using each conditional noncentral hypergeometric distribution for X^*_{1jk}. Originally, their method considered the one–stratum case. However, since the K strata are independent, the generalization of their method to the K–stratum case is not difficult. Let $\chi^*_{1jk}(\psi)$ be the conditional mean of X^*_{1jk}, given the margins n_{1k}, n_{2k}, m^*_{jk}, and $(N_k - m^*_{jk})$. Thus, the difference $X^*_{1jk} - \chi^*_{1jk}(\psi)$ has zero mean. Let $\boldsymbol{X}^* = (X^*_{111}, \ldots, X^*_{1(c-1)K})'$. They constructed an estimating function

$$U(\psi; \boldsymbol{X}^*) = \sum_k \sum_j w^*_{jk}[X^*_{1jk} - \chi^*_{1jk}(\psi)] \,, \tag{2.6}$$

where the weight w^*_{jk} is chosen such that $U(\psi; \boldsymbol{X}^*)$ behaves like a log-likelihood derivative. They claimed that the choice of weights is not critically important and at worst, a poor choice of weights may lead to a small loss of efficiency. For the estimating function (2.6), they proposed a computation method to obtain the estimate and the estimated standard error that will not be discussed here. The estimator produced by the function (2.6) is based on conditional distributions, but not the exact joint conditional distribution. Therefore, we call it a *pseudo conditional likelihood* estimator. Also, they claimed that 'the method of estimation seems to work well and has much in common with quasi–likelihood'.

For the McCullagh and Nelder (1989) method, if any of $\chi_{ijk}(\psi) = \chi^*_{ijk}(\psi) - \chi^*_{i(j-1)k}(\psi)$ is zero for $i = 1, 2$, $j = 1, \ldots, c$ and $k = 1, \ldots, K$, their method fails, since their computation involves the inverse of the cell means. Therefore, the McCullagh and Nelder pseudo conditional likelihood estimator cannot be used for most cases in the sparse–data situation. For the cases which we are interested in, such as a case with one observation in each row, their method is not appropriate because some

conditional cell means will be zero. Their method is good only for the cases in which the sample size is large within each stratum.

The proposed estimator (2.3) is not a model–based estimator; that is, it is not the ML estimator for the common proportional odds model (2.1). We shall show, however, that it behaves better than the ML estimator when the data are sparse. The standard error obtained for the estimator of the common cumulative odds ratio, $\hat{\psi}$, under the naive assumption of independent collapsings is inappropriate. This chapter extends the Robins *et al.* (1986) method and proposes an estimated standard error for the common cumulative odds ratio estimator for an ordinal response case. In Sections 2.3 and 2.4, we show that $\hat{\psi}$ is dually consistent for the common cumulative odds ratio and our proposed estimated standard error for $\hat{\psi}$ is also dually consistent.

2.3 Asymptotics Under the Sparse–Data Limiting Model

In this section, we show that the proposed Mantel–Haenszel–type estimator $\hat{\psi}$ (2.3) is consistent under Model II. Also, we will propose a consistent estimator of the asymptotic variance for the Mantel–Haenszel–type estimator. Under Model II, the number of strata grows with the sample size. In other words, the data within each table are sparse but the number of tables is large.

The proofs of consistency are based on the assumption of a common cumulative odds ratio. For each collapsed table, the cumulative odds ratio is

$$\psi = \frac{\pi^*_{1jk}(1-\pi^*_{2jk})}{\pi^*_{2jk}(1-\pi^*_{1jk})} \ , j \in \{1,\ldots,c-1\} \ , k \in \{1,\ldots,K\} . \tag{2.7}$$

We can write

$$\begin{aligned} \hat{\psi} - \psi &= \frac{\sum_{k=1}^{K}\sum_{j=1}^{c-1}(R_{jk} - \psi S_{jk})}{\sum_{k=1}^{K}\sum_{j=1}^{c-1} S_{jk}} \\ &= \frac{\sum_{k=1}^{K}\sum_{j=1}^{c-1}(R_{jk} - \psi S_{jk})/K}{\sum_{k=1}^{K}\sum_{j=1}^{c-1} S_{jk}/K} . \end{aligned} \tag{2.8}$$

Let Var^a, E^a, Cov^a represent asymptotic variances, asymptotic expectations and asymptotic covariances; let Var, E, Cov represent variances, expectations and covariances.

Under the sampling model of two independent multinomial observations (X_{1jk}, X_{2jk}, $j = 1, \ldots, c$) and by using the assumption of the common cumulative odds ratio (2.7), we have

$$\begin{aligned} \mathrm{E}(R_{jk} - \psi S_{jk}) &= \mathrm{E}(R_{jk}) - \psi \mathrm{E}(S_{jk}) = \frac{n_{1k}\pi^*_{1jk} n_{2k}(1 - \pi^*_{2jk})}{N_k} - \psi \frac{n_{1k}(1 - \pi^*_{1jk}) n_{2k}\pi^*_{2jk}}{N_k} \\ &= \frac{n_{1k} n_{2k}\pi^*_{1jk}(1 - \pi^*_{2jk})}{N_k} - \frac{n_{1k} n_{2k}\pi^*_{1jk}(1 - \pi^*_{2jk})}{N_k} = 0 . \end{aligned}$$

So, if the numerator of $\hat{\psi} - \psi$ in (2.8) minus its mean converges to zero in probability, then the numerator itself converges in probability to zero. Next, we shall apply the Chebyshev weak law of large numbers to show this. The Chebyshev weak law of large numbers applies to sums of non-identically distributed and bounded random variables. It states the following:

Chebyshev weak law of large numbers:

Let $\{Y_n, n \geq 1\}$ be random variables with $E|Y_n|^2 < \infty$, and let $S_n = \sum_{j=1}^{n} Y_j$.

If $\{c_n, n \geq 1\}$ is a sequence of positive constants satisfying $\mathrm{Var}(S_n) = o(c_n^2)$, then

$[S_n - E(S_n)]/c_n \xrightarrow{p} 0$.

Since $\sum_j R_{jk}$ and $\sum_j S_{jk}$ are bounded random variables under Model II, the variance of $\sum_k \sum_j (R_{jk} - \psi S_{jk})$ is $o(K^2)$. Therefore, by the Chebyshev weak law of large numbers, we have

$$\frac{\sum_k \sum_j (R_{jk} - \psi S_{jk}) - \sum_k \sum_j \mathrm{E}(R_{jk} - \psi S_{jk})}{K} \xrightarrow{p} 0 \ , \ \text{as } K \to \infty .$$

Since the expectation equals zero by the previous calculation, we conclude that

$$\frac{\sum_k \sum_j (R_{jk} - \psi S_{jk})}{K} \xrightarrow{p} 0 \ ;$$

that is, the numerator of $\hat{\psi} - \psi$ in (2.8) converges to zero in probability.

Applying the Chebyshev weak law of large numbers again, we can show that the denominator of (2.8) converges in probability to $\lim_{K\to\infty} \sum_{k=1}^{K} \mathrm{E}(\sum_j S_{jk})/K$. This limit is finite and nonzero, under the assumption that ψ is finite and nonzero. Thus, from representation (2.8), $\hat{\psi} - \psi$ converges to zero in probability. Therefore, we conclude that $\hat{\psi}$ is consistent for the common cumulative odds ratio ψ under Model II.

Next, we obtain an asymptotically normal distribution for $\hat{\psi} - \psi$. Since $\sum_j (R_{jk} - \psi S_{jk})$ is a bounded random variable under Model II, we can apply the Liapounov Central Limit Theorem, which states the following:

Liapounov Central Limit Theorem:

If $\{Y_n,\ n \geq 1\}$ are independent random variables with $\mathrm{E}(Y_n) = 0$, $\mathrm{E}(Y_n^2) = \sigma_n^2$, $\mathrm{E}|Y_n|^{2+\delta} < \infty$ for some $\delta > 0$ and all $n \geq 1$, and if $\sum_{j=1}^{n} \mathrm{E}|Y_j|^{2+\delta} = o(s_n^{2+\delta})$, where $s_n^2 = \sum_{j=1}^{n} \sigma_j^2$, $n \geq 1$, then $S_n/s_n \xrightarrow{d} \mathrm{N}(0,1)$, where $S_n = \sum_{j=1}^{n} Y_j$, $n \geq 1$.

Using the Liapounov Central Limit Theorem, we choose $\delta = 1$ and let the random variables $\{Y_n,\ n \geq 1\}$ be $\{\sum_j (R_{jk} - \psi S_{jk})/K,\ k \geq 1\}$. Then, we have $\sum_{k=1}^{K} \mathrm{E}|\sum_j (R_{jk} - \psi S_{jk})/K|^3 = O(1/K^2)$, and $\sum_{k=1}^{K} \mathrm{Var}(\sum_j (R_{jk} - \psi S_{jk})/K) = O(1/K)$. Thus,

$$\frac{\sum_{k=1}^{K} \mathrm{E}|\sum_j (R_{jk} - \psi S_{jk})/K|^3}{(\sum_{k=1}^{K} \mathrm{Var}(\sum_j (R_{jk} - \psi S_{jk})/K))^{3/2}} = \frac{O(1/K^2)}{O(1/K^{3/2})} \to 0 \text{ as } K \to \infty \, .$$

Therefore, the numerator of representation (2.8) for $\hat{\psi} - \psi$ is asymptotically normally distributed. The denominator of (2.8) converges to $\lim_{K\to\infty} \sum_{k=1}^{K} \mathrm{E}(\sum_j S_{jk})/K$ in

probability. From Slutsky's theorem, $\hat{\psi}-\psi$ is also asymptotically normally distributed with

$$\lim_{K\to\infty} K\mathrm{Var}^a(\hat{\psi}) = \frac{\lim_{K\to\infty}\sum_{k=1}^{K}\mathrm{Var}(\sum_j(R_{jk}-\psi S_{jk}))/K}{[\lim_{K\to\infty}\sum_{k=1}^{K}\mathrm{E}(\sum_j S_{jk})/K]^2}$$

$$= \frac{\lim_{K\to\infty}[\sum_k\sum_j \mathrm{Var}(R_{jk}-\psi S_{jk}) + \sum_k\sum_{j<j'} 2\mathrm{Cov}(R_{jk}-\psi S_{jk},\ R_{j'k}-\psi S_{j'k})]/K}{[\lim_{K\to\infty}\sum_k\sum_j \mathrm{E}(S_{jk})/K]^2} \tag{2.9}$$

If we can find a consistent estimator for (2.9), we have a consistent estimator for the asymptotic variance of $\hat{\psi}$ under Model II. We propose a consistent estimator that is the ratio of consistent estimators for estimating the numerator and the denominator of (2.9), which has the form

$$K\hat{\mathrm{V}}\mathrm{ar}(\hat{\psi}) = \frac{\sum_{k=1}^{K}\hat{\xi}_k(\hat{\psi})/K}{[\sum_{k=1}^{K}\sum_{j=1}^{c-1}S_{jk}/K]^2}\ , \tag{2.10}$$

where $\hat{\xi}_k(\hat{\psi}) = \sum_{j=1}^{c-1}\hat{\phi}_{jk}(\hat{\psi}) + 2\sum_{j<j'}^{c-1}\hat{\phi}_{jj'k}(\hat{\psi})$. Here, the term $\hat{\phi}_{jk}(\psi)$ estimates $\mathrm{Var}(R_{jk}-\psi S_{jk})$ and $\hat{\phi}_{jj'k}(\psi)$ estimates $\mathrm{Cov}(R_{jk}-\psi S_{jk}, R_{j'k}-\psi S_{j'k})$ for $j<j'$ in (2.9). The terms $\hat{\phi}_{jk}(\psi)$ and $\hat{\phi}_{jj'k}(\psi)$ are defined as follows:

$$\begin{aligned}\hat{\phi}_{jk}(\psi) &= \frac{n_{1k}n_{2k}}{N_k^2}\{\frac{\psi(n_{1k}-X^*_{1jk})X^*_{2jk}}{n_{1k}}[1+(\psi-1)\frac{X^*_{2jk}}{n_{2k}}]\\ &\quad+\frac{X^*_{1jk}(n_{2k}-X^*_{2jk})}{n_{2k}}[\psi-(\psi-1)\frac{X^*_{1jk}}{n_{1k}}]\}\\ \hat{\phi}_{jj'k}(\psi) &= \frac{n_{1k}n_{2k}}{N_k^2}\{\frac{\psi(n_{1k}-X^*_{1j'k})X^*_{2jk}}{n_{1k}}[1+(\psi-1)\frac{X^*_{2j'k}}{n_{2k}}]\\ &\quad+\frac{X^*_{1jk}(n_{2k}-X^*_{2j'k})}{n_{2k}}[\psi-(\psi-1)\frac{X^*_{1j'k}}{n_{1k}}]\}\ .\end{aligned}$$

The details of consistency are discussed next.

For fixed ψ, since $\{\hat{\xi}_k(\psi), k = 1, \ldots, K\}$ are bounded random variables under Model II, $\mathrm{Var}(\sum_{k=1}^{K} \hat{\xi}_k(\psi)) = o(K^2)$. Applying the Chebyshev weak law of large numbers,

$$\frac{\sum_{k=1}^{K} \hat{\xi}_k(\psi) - \mathrm{E}(\sum_{k=1}^{K} \hat{\xi}_k(\psi))}{K} \xrightarrow{p} 0 ,$$

or

$$\frac{\sum_{k=1}^{K} \hat{\xi}_k(\psi)}{K} \xrightarrow{p} \lim_{K\to\infty} \frac{\sum_{k=1}^{K} \mathrm{E}(\hat{\xi}_k(\psi))}{K} .$$

Therefore, if $\hat{\xi}_k(\psi)$ is an unbiased estimator of $\sum_{j=1}^{c-1} \mathrm{Var}(R_{jk} - \psi S_{jk}) + 2\sum_{j<j'} \mathrm{Cov}(R_{jk} - \psi S_{jk}, R_{j'k} - \psi S_{j'k})$, $\sum_k \hat{\xi}_k(\psi)/K$ will be consistent for estimating the numerator of (2.9) by the weak law of large numbers for summing non-identically distributed, bounded, and unbiased random variables. Also, since $\hat{\psi} \xrightarrow{p} \psi$, and $\sum_k \hat{\xi}_k(\psi)/K$ is a smooth function of ψ, $\sum_k \hat{\xi}_k(\hat{\psi})/K$ will also be consistent by a Taylor series expansion. Furthermore, since S_{jk} is a bounded random variable, we follow the same arguments as above and conclude that

$$\sum_k \sum_j S_{jk}/K \xrightarrow{p} \lim_{K\to\infty} \sum_k \sum_j \mathrm{E}(S_{jk})/K .$$

Thus, the denominator of (2.10) is a consistent estimator for estimating the denominator of (2.9). Next, we will show that $\mathrm{E}(\hat{\phi}_{jk}(\psi)) = \mathrm{Var}(R_{jk} - S_{jk})$ for $j \in \{1, \ldots, c-1\}$ and $k \in \{1, \ldots, K\}$, and $\mathrm{E}(\hat{\phi}_{jj'k}(\psi)) = \mathrm{Cov}(R_{jk} - \psi S_{jk}, R_{j'k} - \psi S_{j'k})$. After we show those expectations, we can claim that the numerator of (2.10) is consistent for estimating the numerator of (2.9), and as a consequence, $K\hat{\mathrm{V}}\mathrm{ar}(\hat{\psi})$ is consistent for (2.9).

For notational convenience, we drop the subscript k while referring to a single $2 \times c$ contingency table. Since $(R_j - \psi S_j)$ can be treated as from a 2×2 contingency table by collapsing the response into a binary response ($\leq j$ and $> j$) for a $2 \times c$ contingency

table, $\mathrm{E}(\hat{\phi}_j(\psi)) = \mathrm{Var}(R_j - \psi S_j)$ follows directly using the same arguments given by Robins *et al.* (1986) for $j \in \{1, \ldots, c-1\}$. We show this first. Define

$$Z_j = (\frac{1}{N^2})(\psi - 1)^2 n_1 \pi^*_{1j}(1 - \pi^*_{1j}) n_2 \pi^*_{2j}(1 - \pi^*_{2j}) , \tag{2.11}$$

$$D_j = \frac{n_1 n_2}{N^2}\{n_2[1 + (\psi - 1)\pi^*_{2j}]\psi\pi^*_{2j}(1 - \pi^*_{1j}) + n_1[\psi - (\psi - 1)\pi^*_{1j}]\pi^*_{1j}(1 - \pi^*_{2j})\} . \tag{2.12}$$

The next lemma shows that the variance of $(R_j - \psi S_j)$ equals $Z_j + D_j$, and the expectation of $\hat{\phi}_j(\psi)$ also equals $Z_j + D_j$. Therefore, $\mathrm{E}(\hat{\phi}_j(\psi)) = \mathrm{Var}(R_j - \psi S_j)$.

Lemma 2.3.1 Assuming the independent multinomial sampling scheme described in Section 2.2 and a common cumulative odds ratio for all collapsings and strata, we have $\mathrm{E}(\hat{\phi}_j(\psi)) = \mathrm{Var}(R_j - \psi S_j)$ for $j \in \{1, \ldots, c-1\}$.

Proof. First, we shall prove that $\mathrm{Var}(R_j - \psi S_j) = Z_j + D_j$. We calculate the variance of $R_j - \psi S_j$ as follows:

$$\begin{aligned}
&\mathrm{Var}(R_j - \psi S_j) = \mathrm{Var}\{(\frac{1}{N})[X^*_{1j}(n_2 - X^*_{2j}) - \psi(n_1 - X^*_{1j})X^*_{2j}]\} \\
= \ &\mathrm{Var}\{(\frac{1}{N})[(\psi - 1)X^*_{1j}X^*_{2j} + n_2 X^*_{1j} - \psi n_1 X^*_{2j}]\} \\
= \ &\frac{1}{N^2}\{(\psi - 1)^2 \mathrm{Var}(X^*_{1j}X^*_{2j}) + n_2^2 \mathrm{Var}(X^*_{1j}) + \psi^2 n_1^2 \mathrm{Var}(X^*_{2j}) \\
&+2(\psi - 1)n_2 \mathrm{Cov}(X^*_{1j}X^*_{2j},\ X^*_{1j}) - 2(\psi - 1)\psi n_1 \mathrm{Cov}(X^*_{1j}X^*_{2j},\ X^*_{2j})\} .
\end{aligned}$$

If X and Y are two independent random variables, $\mathrm{Var}(XY) = \mathrm{Var}(X)\mathrm{Var}(Y) + (\mathrm{E}(X))^2\mathrm{Var}(Y) + (\mathrm{E}(Y))^2\mathrm{Var}(X)$ (Mood *et. al.*, 1973, p. 180). The above equation

equals

$$
\begin{aligned}
&= \frac{1}{N^2}\{(\psi-1)^2[\mathrm{Var}(X^*_{1j})\mathrm{Var}(X^*_{2j}) + (n_1\pi^*_{1j})^2\mathrm{Var}(X^*_{2j}) + (n_2\pi^*_{2j})^2\mathrm{Var}(X^*_{1j})] \\
&\quad +n_2^2 n_1\pi^*_{1j}(1-\pi^*_{1j}) + \psi^2 n_1^2 n_2\pi^*_{2j}(1-\pi^*_{2j}) + 2(\psi-1)n_2[\mathrm{E}((X^*_{1j})^2X^*_{2j}) \\
&\quad -(n_1\pi^*_{1j})^2 n_2\pi^*_{2j}] - 2(\psi-1)\psi n_1[\mathrm{E}(X^*_{1j}(X^*_{2j})^2) - n_1\pi^*_{1j}(n_2\pi^*_{2j})^2]\} \\
&= \frac{1}{N^2}\{(\psi-1)^2[n_1\pi^*_{1j}(1-\pi^*_{1j})n_2\pi^*_{2j}(1-\pi^*_{2j}) + n_1^2(\pi^*_{1j})^2 n_2\pi^*_{2j}(1-\pi^*_{2j}) \\
&\quad +n_2^2(\pi^*_{2j})^2 n_1\pi^*_{1j}(1-\pi^*_{1j})] + n_2^2 n_1\pi^*_{1j}(1-\pi^*_{1j}) + \psi^2 n_1^2 n_2\pi^*_{2j}(1-\pi^*_{2j}) \\
&\quad +2(\psi-1)n_2[(n_1\pi^*_{1j}(1-\pi^*_{1j}) + n_1^2(\pi^*_{1j})^2)n_2\pi^*_{2j} - n_1^2(\pi^*_{1j})^2 n_2\pi^*_{2j}] \\
&\quad -2(\psi-1)\psi n_1[(n_2\pi^*_{2j}(1-\pi^*_{2j}) + n_2^2(\pi^*_{2j})^2)n_1\pi^*_{1j} - n_1\pi^*_{1j}(n_2\pi^*_{2j})^2]\} \\
&= \frac{1}{N^2}(\psi-1)^2 n_1 n_2\pi^*_{1j}(1-\pi^*_{1j})\pi^*_{2j}(1-\pi^*_{2j}) \\
&\quad +\frac{n_1 n_2}{N^2}\{n_2[(\psi-1)^2(\pi^*_{2j})^2\pi^*_{1j}(1-\pi^*_{1j}) + \pi^*_{1j}(1-\pi^*_{1j}) + 2(\psi-1)\pi^*_{1j}(1-\pi^*_{1j})\pi^*_{2j}] \\
&\quad +n_1[(\psi-1)^2(\pi^*_{1j})^2\pi^*_{2j}(1-\pi^*_{2j}) + \psi^2\pi^*_{2j}(1-\pi^*_{2j}) - 2(\psi-1)\psi\pi^*_{1j}\pi^*_{2j}(1-\pi^*_{2j})]\} \\
&= Z_j + \frac{n_1 n_2}{N^2}\{n_2 Q_2 + n_1 Q_1\}\ , \qquad (2.13)
\end{aligned}
$$

where $Q_1 = (\psi-1)^2(\pi^*_{1j})^2\pi^*_{2j}(1-\pi^*_{2j}) + \psi^2\pi^*_{2j}(1-\pi^*_{2j}) - 2(\psi-1)\psi\pi^*_{1j}\pi^*_{2j}(1-\pi^*_{2j})$ and $Q_2 = (\psi-1)^2(\pi^*_{2j})^2\pi^*_{1j}(1-\pi^*_{1j}) + \pi^*_{1j}(1-\pi^*_{1j}) + 2(\psi-1)\pi^*_{1j}(1-\pi^*_{1j})\pi^*_{2j}$.

By the assumption of the common cumulative odds ratio, we have

$$
\psi = \frac{\pi^*_{1j}(1-\pi^*_{2j})}{\pi^*_{2j}(1-\pi^*_{1j})}\ , \quad \psi - 1 = \frac{\pi^*_{1j}-\pi^*_{2j}}{\pi^*_{2j}(1-\pi^*_{1j})}\ ,
$$

for $j \in \{1,\ldots,c-1\}$. Next, using the assumption we shall show that $Q_2 = (1 + (\psi-1)\pi^*_{2j})\psi\pi^*_{2j}(1-\pi^*_{1j})$ and $Q_1 = (\psi-(\psi-1)\pi^*_{1j})\pi^*_{1j}(1-\pi^*_{2j})$, the term (2.13) then

equals $Z_j + D_j$. First, Q_2 can be expressed as

$$\begin{aligned} Q_2 &= (\psi-1)\pi^*_{2j}\pi^*_{1j}(\pi^*_{1j}-\pi^*_{2j}) + (\psi-1)\pi^*_{1j}(1-\pi^*_{1j})\pi^*_{2j} \\ &+ \pi^*_{1j}(1-\pi^*_{1j}) + \pi^*_{1j}(\pi^*_{1j}-\pi^*_{2j}) = (\psi-1)\pi^*_{1j}\pi^*_{2j}(1-\pi^*_{2j}) + \pi^*_{1j}(1-\pi^*_{2j}) \\ &= (1+(\psi-1)\pi^*_{2j})\pi^*_{1j}(1-\pi^*_{2j}) = (1+(\psi-1)\pi^*_{2j})\psi\pi^*_{2j}(1-\pi^*_{1j})\,. \end{aligned}$$

Similarly, we specify Q_1 as

$$\begin{aligned} Q_1 &= \psi^2[(\pi^*_{1j})^2\pi^*_{2j}(1-\pi^*_{2j}) + \pi^*_{2j}(1-\pi^*_{2j}) - 2\pi^*_{1j}\pi^*_{2j}(1-\pi^*_{2j})] \\ &+ \psi[-2(\pi^*_{1j})^2\pi^*_{2j}(1-\pi^*_{2j}) + 2\pi^*_{1j}\pi^*_{2j}(1-\pi^*_{2j})] + (\pi^*_{1j})^2\pi^*_{2j}(1-\pi^*_{2j}) \\ &= \psi^2[\pi^*_{2j}(1-\pi^*_{2j})(1-\pi^*_{1j})^2] + 2\psi[\pi^*_{2j}\pi^*_{1j}(1-\pi^*_{2j})(1-\pi^*_{1j})] + (\pi^*_{1j})^2\pi^*_{2j}(1-\pi^*_{2j}) \\ &= \psi(1-\pi^*_{2j})^2(1-\pi^*_{1j})\pi^*_{1j} + 2(\pi^*_{1j})^2(1-\pi^*_{2j})^2 + (\pi^*_{1j})^2\pi^*_{2j}(1-\pi^*_{2j}) \\ &= \psi(1-\pi^*_{2j})^2(1-\pi^*_{1j})\pi^*_{1j} + (\pi^*_{1j})^2(1-\pi^*_{2j})[2(1-\pi^*_{2j}) + \pi^*_{2j}] \\ &= \psi(1-\pi^*_{2j})^2(1-\pi^*_{1j})\pi^*_{1j} + (\pi^*_{1j})^2(1-\pi^*_{2j})(2-\pi^*_{2j}) \\ &= \psi(1-\pi^*_{2j})^2(1-\pi^*_{1j})\pi^*_{1j} + (\pi^*_{1j})^2(1-\pi^*_{2j})^2 + (\pi^*_{1j})^2(1-\pi^*_{2j}) \\ &= \psi(1-\pi^*_{2j})^2(1-\pi^*_{1j})\pi^*_{1j} + \psi\pi^*_{1j}(1-\pi^*_{2j})\pi^*_{2j}(1-\pi^*_{1j}) + (\pi^*_{1j})^2(1-\pi^*_{2j}) \\ &= \psi\pi^*_{1j}(1-\pi^*_{1j})(1-\pi^*_{2j}) + (\pi^*_{1j})^2(1-\pi^*_{2j}) \\ &= \psi[\pi^*_{1j}(1-\pi^*_{2j}) - (\pi^*_{1j})^2(1-\pi^*_{2j})] + (\pi^*_{1j})^2(1-\pi^*_{2j}) \\ &= \psi\pi^*_{1j}(1-\pi^*_{2j}) - (\psi-1)(\pi^*_{1j})^2(1-\pi^*_{2j}) = (\psi-(\psi-1)\pi^*_{1j})\pi^*_{1j}(1-\pi^*_{2j})\,. \end{aligned}$$

Therefore, $\mathrm{Var}(R_j - \psi S_j) = Z_j + D_j$.

Next, we see that $\mathrm{E}(\hat{\phi}_j(\psi))$ also equals $Z_j + D_j$. The expectation is

$$\begin{aligned}
\mathrm{E}(\hat{\phi}_j(\psi)) &= \frac{n_1 n_2}{N^2}\mathrm{E}\{\frac{\psi(n_1 - X^*_{1j})X^*_{2j}}{n_1}[1 + (\psi - 1)\frac{X^*_{2j}}{n_2}] \\
&\quad + \frac{X^*_{1j}(n_2 - X^*_{2j})}{n_2}[\psi - (\psi - 1)\frac{X^*_{1j}}{n_1}]\} \\
&= \frac{n_1 n_2}{N^2}\{n_2\psi(1 - \pi^*_{1j})\pi^*_{2j}[1 + (\psi - 1)\pi^*_{2j}] + (\psi - 1)\psi(1 - \pi^*_{1j})\pi^*_{2j}(1 - \pi^*_{2j}) \\
&\quad + n_1\pi^*_{1j}(1 - \pi^*_{2j})[\psi - (\psi - 1)\pi^*_{1j}] - (\psi - 1)\pi^*_{1j}(1 - \pi^*_{1j})(1 - \pi^*_{2j})\} \\
&= D_j + \frac{n_1 n_2}{N^2}\{(\psi - 1)\frac{\pi^*_{1j}(1 - \pi^*_{2j})}{\pi^*_{2j}(1 - \pi^*_{1j})}(1 - \pi^*_{1j})\pi^*_{2j}(1 - \pi^*_{2j}) \\
&\quad - (\psi - 1)\pi^*_{1j}(1 - \pi^*_{1j})(1 - \pi^*_{2j})\} \\
&= D_j + \frac{n_1 n_2}{N^2}\{(\psi - 1)\pi^*_{1j}(1 - \pi^*_{2j})^2 - (\psi - 1)\pi^*_{1j}(1 - \pi^*_{1j})(1 - \pi^*_{2j})\} \\
&= D_j + \frac{n_1 n_2}{N^2}\{(\psi - 1)\pi^*_{1j}(1 - \pi^*_{2j})(1 - \pi^*_{2j} - 1 + \pi^*_{1j})\} \\
&= D_j + \frac{n_1 n_2}{N^2}\{(\psi - 1)^2\pi^*_{1j}\pi^*_{2j}(1 - \pi^*_{1j})(1 - \pi^*_{2j})\} = D_j + Z_j \ .
\end{aligned}$$

The above calculation gives that $\mathrm{E}(\hat{\phi}_j(\psi)) = \mathrm{Var}(R_j - \psi S_j)$ for $j \in \{1, \ldots, c-1\}$. □

Now, we consider $\hat{\phi}_{jj'}(\psi)$ and the covariance term $\mathrm{Cov}(R_j - \psi S_j,\ R_{j'} - \psi S_{j'})$ in the variance expression (2.9) for $j < j' \in \{1, \ldots, c-1\}$. The goal is to show that the term $\hat{\phi}_{jj'}(\psi)$ is an unbiased estimator of $\mathrm{Cov}(R_j - \psi S_j,\ R_{j'} - \psi S_{j'})$. In general, the covariance of $R_j - \psi S_j$ and $R_{j'} - \psi S_{j'}$ can be calculated from a 2×3 contingency table formed by collapsing the response into the three-outcomes ($\leq j$, $> j$ and $\leq j'$, $> j'$) for $j < j' \in \{1, \ldots, c-1\}$. For example, in a 2×4 contingency table with cell probabilities $\pi_{11}, \pi_{12}, \pi_{13}, \pi_{14}, \pi_{21}, \pi_{22}, \pi_{23}, \pi_{24}$ satisfying $\sum_{j=1}^{4} \pi_{ij} = 1$ for $i \in \{1, 2\}$, $\mathrm{Cov}(R_1 - \psi S_1,\ R_3 - \psi S_3)$ can be calculated from a 2×3 contingency table collapsing the four response categories into three response categories (1, 2 and 3, 4). The new 2×3 contingency table has cell probabilities (counts) $\pi'_{i1} = \pi_{i1}$

$(X'_{i1} = X_{i1})$, $\pi'_{i2} = \pi_{i2} + \pi_{i3}$ $(X'_{i2} = X_{i2} + X_{i3})$, and $\pi'_{i3} = \pi_{i4}$ $(X'_{i3} = X_{i4})$ for $i \in \{1, 2\}$. So, $\mathrm{Cov}(R_1 - \psi S_1,\ R_3 - \psi S_3)$ for the 2×4 contingency table is equivalent to $\mathrm{Cov}(R'_1 - \psi S'_1,\ R'_2 - \psi S'_2)$ for the 2×3 contingency table, where R_j $(= X^*_{1j}(n_2 - X^*_{2j})/N)$ and S_j $(= (n_1 - X^*_{1j})X^*_{2j}/N)$ correspond to the original table, whereas R'_j $(= X'^*_{1j}(n_2 - X'^*_{2j})/N)$ and S'_j $(= (n_1 - X'^*_{1j})X'^*_{2j}/N)$ correspond to the new 2×3 contingency table.

Because of this covariance property, we consider a $2 \times 3 \times K$ case first. Then, the generalization to a general $2 \times c \times K$ case follows easily. Similarly, for convenience, we drop the subscript k, referring to a single 2×3 contingency table. In a 2×3 contingency table, we consider $\hat{\phi}_{12}(\psi)$ and the covariance term $\mathrm{Cov}(R_1 - \psi S_1,\ R_2 - \psi S_2)$ in the variance expression (2.9). Define

$$Z_{12} = (\frac{1}{N^2})(\psi - 1)^2 n_1 n_2 \pi_{11} \pi_{13} \pi_{21} \pi_{23}$$

and

$$D_{12} = \frac{n_1 n_2}{N^2}\{n_2[1 + (\psi - 1)(1 - \pi_{23})]\psi \pi_{13}\pi_{21} + n_1[\psi - (\psi - 1)(1 - \pi_{13})]\pi_{11}\pi_{23}\}.$$

The next lemma shows that the covariance of $R_1 - \psi S_1$, and $R_2 - \psi S_2$ equals $Z_{12} + D_{12}$, and the expectation of $\hat{\phi}_{12}(\psi)$ also equals $Z_{12} + D_{12}$. Then, we have $\mathrm{Cov}(R_1 - \psi S_1,\ R_2 - \psi S_2) = \hat{\phi}_{12}(\psi)$.

Lemma 2.3.2 Assuming the independent multinomial sampling scheme described in Section 2.2 and a common cumulative odds ratio for all collapsings and strata, we have $\mathrm{E}(\hat{\phi}_{12}(\psi)) = \mathrm{Cov}(R_1 - \psi S_1,\ R_2 - \psi S_2)$

Proof. Under the common cumulative odds ratio assumption, we have

$$\psi = \frac{\pi_{11}(1 - \pi_{21})}{\pi_{21}(1 - \pi_{11})} = \frac{(1 - \pi_{13})\pi_{23}}{(1 - \pi_{23})\pi_{13}}, \tag{2.14}$$

$$\psi - 1 = \frac{\pi_{11} - \pi_{21}}{\pi_{21}(1 - \pi_{11})} = \frac{\pi_{23} - \pi_{13}}{(1 - \pi_{23})\pi_{13}}. \tag{2.15}$$

Also, notice that

$$\mathrm{Cov}(X_{1h},\, X_{1h'}X_{2j}) = -n_1 n_2 \pi_{1h}\pi_{1h'}\pi_{2j}\ ,$$

$$\mathrm{Cov}(X_{2j},\, X_{1h}X_{2j'}) = -n_1 n_2 \pi_{1h}\pi_{2j}\pi_{2j'}\ ,$$

$$\mathrm{Cov}(X_{1h}X_{2j},\, X_{1h'}X_{2j'}) = n_1 n_2 \pi_{1h}\pi_{1h'}\pi_{2j}\pi_{2j'}(-n_2 - n_1 + 1),$$

for $h \neq h' \in \{1,2,3\}$ and $j \neq j' \in \{1,2,3\}$. We will use those equalities in the following proof. First, we specify

$$\begin{aligned}
&\mathrm{Cov}(R_1 - \psi S_1,\ R_2 - \psi S_2) \\
=\ & \frac{1}{N^2}\{\mathrm{Cov}[X_{11}(n_2 - X_{21}) - \psi(n_1 - X_{11})X_{21}\ ,\ (n_1 - X_{13})X_{23} - \psi X_{13}(n_2 - X_{23})]\} \\
=\ & \frac{1}{N^2}\{\mathrm{Cov}[n_2 X_{11} - n_1\psi X_{21} + (\psi - 1)X_{11}X_{21}, n_1 X_{23} - n_2\psi X_{13} + (\psi - 1)X_{13}X_{23}]\} \\
=\ & \frac{1}{N^2}\{-n_2^2\psi\mathrm{Cov}(X_{11}, X_{13}) + (\psi - 1)n_2\mathrm{Cov}(X_{11}, X_{13}X_{23}) - n_1^2\psi\mathrm{Cov}(X_{21}, X_{23}) \\
& -n_1\psi(\psi - 1)\mathrm{Cov}(X_{21}\ ,\ X_{13}X_{23}) + (\psi - 1)n_1\mathrm{Cov}(X_{11}X_{21}\ ,\ X_{23}) \\
& -n_2\psi(\psi - 1)\mathrm{Cov}(X_{11}X_{21}\ ,\ X_{13}) + (\psi - 1)^2\mathrm{Cov}(X_{11}X_{21}\ ,\ X_{13}X_{23})\} \\
=\ & \frac{1}{N^2}\{n_2^2\psi n_1\pi_{11}\pi_{13} - (\psi - 1)n_2^2 n_1\pi_{11}\pi_{13}\pi_{23} + n_1^2 n_2\psi\pi_{21}\pi_{23} \\
& +n_1^2\psi(\psi - 1)n_2\pi_{21}\pi_{13}\pi_{23} - n_1^2 n_2(\psi - 1)\pi_{11}\pi_{21}\pi_{23} \\
& +n_2^2 n_1\psi(\psi - 1)\pi_{11}\pi_{21}\pi_{13} + (\psi - 1)^2 n_2 n_1\pi_{11}\pi_{13}\pi_{21}\pi_{23}(-n_2 - n_1 + 1)\} \\
=\ & \frac{n_1 n_2}{N^2}\{[(\psi - 1)^2\pi_{11}\pi_{13}\pi_{21}\pi_{23}] \\
& +n_2[\psi\pi_{11}\pi_{13} - (\psi - 1)\pi_{11}\pi_{13}\pi_{23} + \psi(\psi - 1)\pi_{11}\pi_{21}\pi_{13} - (\psi - 1)^2\pi_{11}\pi_{13}\pi_{21}\pi_{23}] \\
& +n_1[\psi\pi_{21}\pi_{23} + \psi(\psi - 1)\pi_{21}\pi_{13}\pi_{23} - (\psi - 1)\pi_{11}\pi_{21}\pi_{23} - (\psi - 1)^2\pi_{11}\pi_{13}\pi_{21}\pi_{23}]\} \\
=\ & Z_{12} + \frac{n_1 n_2}{N^2}[n_2 Q_2' + n_1 Q_1']\ ,
\end{aligned} \tag{2.16}$$

where $Q_2' = \psi\pi_{11}\pi_{13} - (\psi-1)\pi_{11}\pi_{13}\pi_{23} + \psi(\psi-1)\pi_{11}\pi_{21}\pi_{13} - (\psi-1)^2\pi_{11}\pi_{13}\pi_{21}\pi_{23}$ and $Q_1' = \psi\pi_{21}\pi_{23} + \psi(\psi-1)\pi_{21}\pi_{13}\pi_{23} - (\psi-1)\pi_{11}\pi_{21}\pi_{23} - (\psi-1)^2\pi_{11}\pi_{13}\pi_{21}\pi_{23}$.

By the common cumulative odds ratio assumption (2.14) and (2.15), we will show next that the terms Q_2' and Q_1' can simplify to $[1+(\psi-1)(1-\pi_{23})]\psi\pi_{13}\pi_{21}$ and $[\psi-(\psi-1)(1-\pi_{13})]\pi_{11}\pi_{23}$, respectively. Then, the covariance (2.16) equals $Z_{12}+D_{12}$. First, we specify Q_2' as

$$\begin{aligned} Q_2' &= \pi_{11}\pi_{13}\pi_{23} - \pi_{11}\pi_{13}\pi_{21}\pi_{23} + \psi[\pi_{11}\pi_{13} - \pi_{11}\pi_{13}\pi_{23} - \pi_{11}\pi_{21}\pi_{13} + 2\pi_{11}\pi_{13}\pi_{21}\pi_{23}] \\ &\quad +\psi^2[\pi_{11}\pi_{21}\pi_{13} - \pi_{11}\pi_{13}\pi_{21}\pi_{23}] \\ &= \pi_{11}\pi_{13}\pi_{23}(1-\pi_{21}) + \psi\pi_{11}\pi_{13}(1-\pi_{23}) + \psi\pi_{11}\pi_{21}\pi_{13}(\pi_{23}-1) \\ &\quad +\psi\pi_{11}\pi_{13}\pi_{21}\pi_{23} + \psi^2\pi_{11}\pi_{13}\pi_{21}(1-\pi_{23}). \end{aligned}$$

By (2.14), this equals

$$\begin{aligned} &\pi_{11}\pi_{13}\pi_{23}(1-\pi_{21}) + \pi_{11}(1-\pi_{13})\pi_{23} - \pi_{11}\pi_{21}\pi_{23}(1-\pi_{13}) \\ &+\psi\pi_{11}\pi_{13}\pi_{21}\pi_{23} + \psi\pi_{11}\pi_{21}(1-\pi_{13})\pi_{23} \\ =\ &\pi_{11}\pi_{13}\pi_{23} - \pi_{11}\pi_{13}\pi_{21}\pi_{23} + \pi_{11}\pi_{23} - \pi_{11}\pi_{13}\pi_{23} - \pi_{11}\pi_{21}\pi_{23} + \pi_{11}\pi_{21}\pi_{13}\pi_{23} \\ &+\psi\pi_{11}\pi_{13}\pi_{21}\pi_{23} + \psi\pi_{11}\pi_{21}\pi_{23} - \psi\pi_{11}\pi_{21}\pi_{13}\pi_{23} \\ =\ &\pi_{11}\pi_{23} - \pi_{11}\pi_{21}\pi_{23} + \psi\pi_{11}\pi_{21}\pi_{23} = \pi_{11}\pi_{23}(1-\pi_{21}) + \psi\pi_{11}\pi_{21}\pi_{23}. \end{aligned}$$

By (2.14) again, this equals

$$\psi\pi_{23}\pi_{21}(1-\pi_{11}) + \psi\pi_{11}\pi_{21}\pi_{23} = \psi\pi_{21}\pi_{23} = \psi\pi_{13}\pi_{21} + \psi(\pi_{23}-\pi_{13})\pi_{21}.$$

Finally, by (2.15), this equals

$$\psi\pi_{13}\pi_{21} + (\psi-1)(1-\pi_{23})\psi\pi_{13}\pi_{21} = [1+(\psi-1)(1-\pi_{23})]\psi\pi_{13}\pi_{21}.$$

Similarly, the term Q_1' equals

$$\begin{aligned} Q_1' &= \pi_{11}\pi_{21}\pi_{23} - \pi_{11}\pi_{13}\pi_{21}\pi_{23} + \psi[\pi_{21}\pi_{23} - \pi_{21}\pi_{13}\pi_{23} - \pi_{11}\pi_{21}\pi_{23} + 2\pi_{11}\pi_{13}\pi_{21}\pi_{23}] \\ &\quad +\psi^2[\pi_{21}\pi_{13}\pi_{23} - \pi_{11}\pi_{13}\pi_{21}\pi_{23}] \\ &= \pi_{11}\pi_{21}\pi_{23}(1-\pi_{13}) + \psi\pi_{21}\pi_{23}(1-\pi_{13}) + \psi\pi_{11}\pi_{21}\pi_{23}(\pi_{13}-1) \\ &\quad +\psi\pi_{11}\pi_{13}\pi_{21}\pi_{23} + \psi^2\pi_{21}\pi_{13}\pi_{23}(1-\pi_{11})\,. \end{aligned}$$

By (2.14), this equals

$$\begin{aligned} &\psi\pi_{11}\pi_{21}\pi_{13}(1-\pi_{23}) + \psi^2\pi_{21}(1-\pi_{23})\pi_{13} - \psi^2\pi_{11}\pi_{21}(1-\pi_{23})\pi_{13} \\ &\quad +\psi\pi_{11}\pi_{13}\pi_{21}\pi_{23} + \psi^2\pi_{21}\pi_{13}\pi_{23}(1-\pi_{11}) \\ &= \psi\pi_{11}\pi_{21}\pi_{13} - \psi\pi_{11}\pi_{21}\pi_{13}\pi_{23} + \psi^2\pi_{21}\pi_{13} - \psi^2\pi_{21}\pi_{23}\pi_{13} - \psi^2\pi_{11}\pi_{21}\pi_{13} \\ &\quad +\psi^2\pi_{11}\pi_{21}\pi_{13}\pi_{23} + \psi\pi_{11}\pi_{13}\pi_{21}\pi_{23} + \psi^2\pi_{21}\pi_{13}\pi_{23} - \psi^2\pi_{21}\pi_{13}\pi_{23}\pi_{11} \\ &= \psi\pi_{11}\pi_{21}\pi_{13} + \psi^2\pi_{21}\pi_{13} - \psi^2\pi_{11}\pi_{21}\pi_{13} = \psi\pi_{11}\pi_{21}\pi_{13} + \psi^2\pi_{21}\pi_{13}(1-\pi_{11})\,. \end{aligned}$$

By (2.14) again, this equals

$$\psi\pi_{11}\pi_{21}\pi_{13} + \psi\pi_{13}\pi_{11}(1-\pi_{21}) = \psi\pi_{11}\pi_{13} = \psi\pi_{11}\pi_{23} - (\pi_{23}-\pi_{13})\psi\pi_{11}\,.$$

By (2.15), this equals

$$\begin{aligned} &\pi_{11}\pi_{23}\psi - (\psi-1)\psi(1-\pi_{23})\pi_{13}\pi_{11} \\ &= \pi_{11}\pi_{23}\psi - (\psi-1)(1-\pi_{13})\pi_{11}\pi_{23} = [\psi - (\psi-1)(1-\pi_{13})]\pi_{11}\pi_{23}\,. \end{aligned}$$

Therefore, the covariance (2.16) equals $Z_{12} + D_{12}$.

Next, we show that $E(\hat{\phi}_{12}(\psi)) = Z_{12} + D_{12}$. The expectation is

$$
\begin{aligned}
E(\hat{\phi}_{12}(\psi)) &= \frac{n_1 n_2}{N^2} E\{\frac{\psi(n_1 - X^*_{12})X_{21}}{n_1}[1 + (\psi - 1)\frac{X^*_{22}}{n_2}] \\
&+ \frac{X_{11}(n_2 - X^*_{22})}{n_2}[\psi - (\psi - 1)\frac{X^*_{12}}{n_1}]\} \\
&= \frac{n_1 n_2}{N^2} E\{\frac{\psi X_{13} X_{21}}{n_1}[1 + (\psi - 1)(1 - \frac{X_{23}}{n_2})] + \frac{X_{11}X_{23}}{n_2}[\psi - (\psi - 1)(1 - \frac{X_{13}}{n_1})]\} \\
&= \frac{n_1 n_2}{N^2}\{n_2 \pi_{13}\pi_{21}[1 + (\psi - 1)(1 - \pi_{23})] + n_1 \pi_{11}\pi_{23}[\psi - (\psi - 1)(1 - \pi_{13})] \\
&\quad - \frac{\psi E(X_{13})}{n_1 n_2}(\psi - 1)\mathrm{Cov}(X_{21}, X_{23}) + \frac{E(X_{23})}{n_1 n_2}(\psi - 1)\mathrm{Cov}(X_{11}, X_{13})\} \\
&= D_{12} + \frac{n_1 n_2}{N^2}[\psi \pi_{13}(\psi - 1)\pi_{21}\pi_{23} - (\psi - 1)\pi_{23}\pi_{11}\pi_{13}] \\
&= D_{12} + \frac{n_1 n_2}{N^2}[(\psi - 1)\pi_{13}\pi_{23}(\psi \pi_{21} - \pi_{11})] \\
&= D_{12} + \frac{n_1 n_2}{N^2}\{(\psi - 1)^2 \pi_{13}\pi_{23}[\frac{\pi_{21}(1 - \pi_{11})}{\pi_{11} - \pi_{21}}][\frac{\pi_{11}(1 - \pi_{21})}{1 - \pi_{11}} - \pi_{11}]\} \\
&= D_{12} + \frac{n_1 n_2}{N^2}(\psi - 1)^2 \pi_{13}\pi_{23}[\frac{\pi_{21}(1 - \pi_{11})}{\pi_{11} - \pi_{21}}][\frac{\pi_{11}(\pi_{11} - \pi_{21})}{1 - \pi_{11}}] = D_{12} + Z_{12}\,.
\end{aligned}
$$

Since $\mathrm{Cov}(R_1 - \psi S_1,\ R_2 - \psi S_2) = D_{12} + Z_{12} = E(\hat{\phi}_{12}(\psi))$, we have shown that $E(\hat{\phi}_{12}(\psi)) = \mathrm{Cov}(R_1 - \psi S_1,\ R_2 - \psi S_2)$. □

For the general $2 \times c \times K$ case, Lemma 2.3.2 can be generalized to $E(\hat{\phi}_{jj'k}(\psi)) = \mathrm{Cov}(R_{jk} - \psi S_{jk},\ R_{j'k} - \psi S_{j'k})$ for $j < j' \in \{1, \ldots, c-1\}$ and $k \in \{1, \ldots, K\}$. Let $\underline{X}^*_{ijk} = X_{i(j+1)k} + \cdots + X_{ick}$ and $\underline{\pi}^*_{ijk} = \pi_{i(j+1)k} + \cdots + \pi_{ick}$. Again, we drop the subscript k for simplicity. Define

$$
Z_{jj'} = (\frac{1}{N^2})(\psi - 1)^2 n_1 n_2 \pi^*_{1j}\underline{\pi}^*_{1j'}\pi^*_{2j}\underline{\pi}^*_{2j'} \tag{2.17}
$$

and

$$
D_{jj'} = \frac{n_1 n_2}{N^2}\{n_2[1 + (\psi - 1)(1 - \underline{\pi}^*_{2j'})]\psi \underline{\pi}^*_{1j'}\pi^*_{2j} + n_1[\psi - (\psi - 1)(1 - \underline{\pi}^*_{1j'})]\pi^*_{1j}\underline{\pi}^*_{2j'}\}\,. \tag{2.18}
$$

In the proof of the Lemma 2.3.2, we replace all X_{i1}, X_{i3}, π_{i1}, and π_{i3}, by X^*_{ij}, $X^*_{ij'}$, π^*_{ij}, $\pi^*_{ij'}$, respectively for $i \in \{1,2\}$, and $j < j' \in \{1,\ldots,c-1\}$. Then, following the same procedures, we shall show that the covariance of $R_j - \psi S_j$, and $R_{j'} - \psi S_{j'}$ equals $Z_{jj'} + D_{jj'}$, and the expectation of $\hat{\phi}_{jj'}(\psi)$ also equals $Z_{jj'} + D_{jj'}$ for $j < j' \in \{1,\ldots,c-1\}$. Thus, we have $\mathrm{Cov}(R_{jk} - \psi S_{jk},\ R_{j'k} - \psi S_{j'k}) = \hat{\phi}_{jj'k}(\psi)$.

From Lemma 2.3.1 and Lemma 2.3.2, we conclude that $\hat{\xi}_k(\psi)$ is an unbiased estimator of $\sum_{j=1}^{c-1}\mathrm{Var}(R_{jk} - \psi S_{jk}) + 2\sum_{j<j'}^{c-1}\mathrm{Cov}(R_{jk} - \psi S_{jk},\ R_{j'k} - \psi S_{j'k})$. Therefore, $K\hat{\mathrm{Var}}(\hat{\psi})$ as expressed in (2.10) is consistent for the asymptotic variance (2.9) under Model II (sparse-data).

2.4 Asymptotics Under the Large–Stratum Limiting Model

We next discuss consistency for our proposed estimator and the proposed estimated variance under Model I. Under Model I, the number of tables K is fixed and the sample size for each table increases indefinitely (all $N_k \to \infty$). We first show that our proposed Mantel–Haenszel–type estimator (2.3) is consistent for the common cumulative odds ratio under the large–stratum limiting model. After that, we obtain an asymptotic variance of our estimator under that limiting model, and then we show that the estimator obtained in the previous section is consistent for it.

Hauck (1979) proposed a large sample variance of the ordinary Mantel–Haenszel estimator for $2 \times 2 \times K$ contingency tables. He derived an asymptotic distribution of $\hat{\psi}_{\mathrm{MH}}$ for a general case in which the individual odds ratios $\psi_1,\ldots,\psi_K$ need not be equal. Furthermore, based on this result, he provided an asymptotic distribution of $\hat{\psi}_{\mathrm{MH}}$ when $\psi_1 = \cdots = \psi_K$. Guilbaud (1983) noted and corrected an error of Hauck's procedure in the general case. However, Guilbaud's asymptotic variance is in accordance with Hauck's asymptotic variance when $\psi_1 = \cdots = \psi_K$. Here, we generalize Guilbaud's approach for use with our proposed estimator for $2 \times c \times K$ contingency

tables. Also, the asymptotic distribution of $\hat{\psi}$ from Guilbaud's approach can simplify to the asymptotic distribution from Hauck's approach under the assumption of the common cumulative odds ratio. Calculations for both Guilbaud's and Hauck's approaches are based on the delta method to obtain an asymptotic variance under Model I.

For a $2 \times c \times K$ contingency table, denote the cumulative odds ratio estimator for the kth stratum by

$$\hat{\psi}_k = \frac{\sum_{j=1}^{c-1} X^*_{1jk}(n_{2k} - X^*_{2jk})}{\sum_{j=1}^{c-1} X^*_{2jk}(n_{1k} - X^*_{1jk})},$$

for $k \in \{1, \ldots, K\}$. Also, let

$$\hat{t}_k = (\frac{1}{n_{1k}} + \frac{1}{n_{2k}})^{-1}[\sum_{j=1}^{c-1}(1 - p^*_{1jk})p^*_{2jk}],$$

where $p^*_{1jk} = X^*_{1jk}/n_{1k}$ and $p^*_{2jk} = X^*_{2jk}/n_{2k}$ for $k \in \{1, \ldots, K\}$. The term $\hat{t}_k$ can be expressed as

$$\begin{aligned} \hat{t}_k &= (\frac{n_{1k} + n_{2k}}{n_{1k}n_{2k}})^{-1}[\sum_{j=1}^{c-1}(1 - \frac{X^*_{1jk}}{n_{1k}})(\frac{X^*_{2jk}}{n_{2k}})] \\ &= \frac{1}{n_{1k} + n_{2k}}[\sum_{j=1}^{c-1}(n_{1k} - X^*_{1jk})X^*_{2jk}] = \sum_j S_{jk}, \end{aligned}$$

and $\hat{t}_k\hat{\psi}_k$ can be expressed as

$$\begin{aligned} \hat{t}_k\hat{\psi}_k &= \frac{1}{n_{1k} + n_{2k}}[\sum_{j=1}^{c-1}(n_{1k} - X^*_{1jk})X^*_{2jk}]\frac{\sum_j X^*_{1jk}(n_{2k} - X^*_{2jk})}{\sum_j X^*_{2jk}(n_{1k} - X^*_{1jk})} \\ &= \frac{1}{n_{1k} + n_{2k}}[\sum_j X^*_{1jk}(n_{2k} - X^*_{2jk})] = \sum_j R_{jk}. \end{aligned}$$

Thus, we can re-write $\hat{\psi}$ as

$$\hat{\psi} = \frac{\sum_k \hat{t}_k\hat{\psi}_k}{\sum_k \hat{t}_k} = \sum_k \hat{w}_k\hat{\psi}_k, \tag{2.19}$$

where $\hat{w}_k = \hat{t}_k / \sum_k \hat{t}_k$.

Since the number of tables is fixed and the sample size for each table increases without bound while $N_+a_{1k} = n_{1k}$ and $N_+a_{2k} = n_{2k}$, we consider a limiting distribution for each $\hat{\psi}_k$. By the weak law of large numbers,

$$\hat{\psi}_k = \frac{\sum_{j=1}^{c-1} \frac{X^*_{1jk}}{n_{1k}} \frac{(n_{2k} - X^*_{2jk})}{n_{2k}}}{\sum_{j=1}^{c-1} \frac{X^*_{2jk}}{n_{2k}} \frac{(n_{1k} - X^*_{1jk})}{n_{1k}}} \xrightarrow{p} \frac{\sum_{j=1}^{c-1} \pi^*_{1jk}(1 - \pi^*_{2jk})}{\sum_{j=1}^{c-1} \pi^*_{2jk}(1 - \pi^*_{1jk})} = \psi_k .$$

Similarly,

$$\hat{w}_k = \hat{t}_k / \sum_k \hat{t}_k \xrightarrow{p} t_k / \sum_k t_k = w_k ,$$

where $t_k = (a_{1k}^{-1} + a_{2k}^{-1})^{-1} [\sum_{j=1}^{c-1} (1 - \pi^*_{1jk}) \pi^*_{2jk}]$, and as a consequence,

$$\hat{\psi} \xrightarrow{p} \sum_k w_k \psi_k .$$

Under the common cumulative odds ratio assumption, we have $\psi_k = \psi$. Therefore, $\hat{\psi}$ converges to ψ in probability under Model I.

Alternatively, when we consider the odds ratio for each collapsing and each stratum, the form of $\hat{\psi}$ (2.19) also can be expressed as

$$\hat{\psi} = \frac{\sum_k \sum_j \hat{t}_{jk} \hat{\psi}_{jk}}{\sum_k \sum_j \hat{t}_{jk}} = \sum_k \sum_j \hat{w}_{jk} \hat{\psi}_{jk} , \tag{2.20}$$

where $\hat{w}_{jk} = \hat{t}_{jk} / \sum_k \sum_j \hat{t}_{jk}$. In this expression (2.20),

$$\hat{t}_{jk} = (\frac{1}{n_{1k}} + \frac{1}{n_{2k}})^{-1} [(1 - p^*_{1jk}) p^*_{2jk}] ,$$

and

$$\hat{\psi}_{jk} = \frac{X^*_{1jk}(n_{2k} - X^*_{2jk})}{X^*_{2jk}(n_{1k} - X^*_{1jk})} .$$

Under the common cumulative odds ratio assumption, we have constant odds ratio not only across the K strata but also across the $c - 1$ collapsings; that is, $\psi_{jk} =$

ψ. Therefore, under the assumption, $\hat{\psi}$ is consistent for ψ under Model I following the same arguments as the expression (2.19). However, two expressions (2.19) and (2.20) are equivalent. Next, we obtain an asymptotic distribution only based on the expression (2.19). The asymptotic distribution derived is for a general case without assuming the common odds ratio across the strata or the collapsings.

Since $\sqrt{N_+}(\hat{\psi}_k - \psi_k)$ and $\sqrt{N_+}(\hat{w}_k - w_k)$ are asymptotically linear in $\sqrt{N_+}(X_{ijk}/n_{ik} - \pi_{ijk})$, we will use the delta method to present an asymptotic distribution for our estimator in Lemma 2.4.1. The error that Guilbaud noted in Hauck's derivation lies in the fact that the sampling errors of the weights $\hat{w}_k$ must be taken into account.

Lemma 2.4.1 As $N_+ \to \infty$ with $N_+ a_{ik} = n_{ik}$, where $a_{ik} \neq 0$ for $i \in \{1,2\}$ and $k \in \{1,\ldots,K\}$, $\sqrt{N_+}(\hat{\psi} - \psi)$ is asymptotically normal with mean 0 and variance

$$
\begin{aligned}
V_{\hat{\psi}} \;=\; & \frac{1}{(\sum_{l=1}^{K} t_l)^2}\{\sum_{k=1}^{K}\frac{t_k^2}{a_{1k}}[\sum_{a=1}^{c}\pi_{1ak}(\frac{\psi_k}{A_k}\sum_{b=a+1}^{c}(b-a)\pi_{2bk} - \frac{\psi}{B_k}\sum_{b=1}^{a-1}(a-b)\pi_{2bk})^2] \\
& + \sum_{k=1}^{K}\frac{t_k^2}{a_{2k}}[\sum_{b=1}^{c}\pi_{2bk}(\frac{\psi_k}{A_k}\sum_{a=1}^{b-1}(b-a)\pi_{1ak} - \frac{\psi}{B_k}\sum_{a=b+1}^{c}(a-b)\pi_{1ak})^2]\} \\
& - \sum_{k}(a_{1k}^{-1} + a_{2k}^{-1})[w_k(\psi_k - \psi)]^2 ,
\end{aligned}
$$

where $A_k = \sum_{a<b}^{c}(b-a)\pi_{1ak}\pi_{2bk} = \sum_j \pi^*_{1jk}(1-\pi^*_{2jk})$, $B_k = \sum_{a>b}^{c}(a-b)\pi_{1ak}\pi_{2bk}$ $= \sum_j \pi^*_{2jk}(1-\pi^*_{1jk})$, $a \in \{1,\ldots,c\}$, $b \in \{1,\ldots,c\}$, and $k \in \{1,\ldots,K\}$.

Proof. We can express ψ as

$$\begin{aligned}
\psi &= \sum_k w_k\psi_k = \sum_k \exp[\log(w_k\psi_k)] \\
&= \sum_k \exp[\log t_k - \log\sum_{l=1}^{K} t_l + \log(\sum_{j=1}^{c-1}\pi^*_{1jk}(1-\pi^*_{2jk})) - \log(\sum_{j=1}^{c-1}\pi^*_{2jk}(1-\pi^*_{1jk}))] \\
&= \sum_k \exp\{\log(a_{1k}^{-1}+a_{2k}^{-1})^{-1} + \log(\sum_{a>b}^{c}(a-b)\pi_{1ak}\pi_{2bk}) - \log[\sum_{l=1}^{K}(a_{1l}^{-1}+a_{2l}^{-1})^{-1} \\
&\quad (\sum_{a>b}^{c}(a-b)\pi_{1al}\pi_{2bl})] + \log(\sum_{a<b}^{c}(b-a)\pi_{1ak}\pi_{2bk}) - \log(\sum_{a>b}^{c}(a-b)\pi_{1ak}\pi_{2bk})\} \\
&= \sum_k \exp\{\log(a_{1k}^{-1}+a_{2k}^{-1})^{-1} + \log(A_k) - \log[\sum_{l=1}^{K}(a_{1l}^{-1}+a_{2l}^{-1})^{-1}B_l]\}\,.
\end{aligned}$$

We differentiate ψ with respect to π_{1ak} and π_{2bk}, respectively. We have

$$\begin{aligned}
\zeta_{1ak} &= \frac{\partial\psi}{\partial\pi_{1ak}} = w_k\psi_k[\frac{1}{A_k}\sum_{b=a+1}^{c}(b-a)\pi_{2bk} - \frac{1}{\sum_l t_l}(a_{1k}^{-1}+a_{2k}^{-1})^{-1}(\sum_{b=1}^{a-1}(a-b)\pi_{2bk})] \\
&\quad - \sum_{k'\neq k} w_{k'}\psi_{k'}[\frac{1}{\sum_l t_l}(a_{1k}^{-1}+a_{2k}^{-1})^{-1}(\sum_{b=1}^{a-1}(a-b)\pi_{2bk})]\,.
\end{aligned}$$

Since $\sum_{k'\neq k} w_{k'}\psi_{k'} = \psi - w_k\psi_k$,

$$\zeta_{1ak} = \frac{t_k\psi_k}{\sum_l t_l}[\frac{1}{A_k}\sum_{b=a+1}^{c}(b-a)\pi_{2bk}] - \frac{\psi}{\sum_l t_l}(a_{1k}^{-1}+a_{2k}^{-1})^{-1}(\sum_{b=1}^{a-1}(a-b)\pi_{2bk})\,.$$

Similarly,

$$\begin{aligned}
\zeta_{2bk} &= \frac{\partial\psi}{\partial\pi_{2bk}} = w_k\psi_k[\frac{1}{A_k}\sum_{a=1}^{b-1}(b-a)\pi_{1ak} - \frac{1}{\sum_l t_l}(a_{1k}^{-1}+a_{2k}^{-1})^{-1}(\sum_{a=b+1}^{c}(a-b)\pi_{1ak})] \\
&\quad - \sum_{k'\neq k} w_{k'}\psi_{k'}[\frac{1}{\sum_l t_l}(a_{1k}^{-1}+a_{2k}^{-1})^{-1}(\sum_{a=b+1}^{c}(a-b)\pi_{1ak})] \\
&= \frac{t_k\psi_k}{\sum_l t_l}[\frac{1}{A_k}\sum_{a=1}^{b-1}(b-a)\pi_{1ak}] - \frac{\psi}{\sum_l t_l}(a_{1k}^{-1}+a_{2k}^{-1})^{-1}(\sum_{a=b+1}^{c}(a-b)\pi_{1ak})\,.
\end{aligned}$$

By the Central Limit Theorem, the sample proportions from the multinomial distributions are asymptotically multivariate normally distributed. Then,

$$\sqrt{N_+}(\boldsymbol{p}-\boldsymbol{\pi}) \xrightarrow{d} N\left(\mathbf{0}, \Sigma = \begin{bmatrix} \Sigma_1 & 0 & 0 & \dots & 0 \\ 0 & \Sigma_2 & 0 & \dots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \dots & \Sigma_K \end{bmatrix}\right),$$

where $\boldsymbol{p} = (\boldsymbol{p}'_1, \dots, \boldsymbol{p}'_K)'$, $\boldsymbol{p}_k = (\boldsymbol{p}'_{1k}\ \boldsymbol{p}'_{2k})'$, $\boldsymbol{p}_{ik} = (p_{i1k}, p_{i2k}, \dots, p_{ick})'$, with analogous notations for $\boldsymbol{\pi}$. The variance and covariance matrix is

$$\Sigma_k = \begin{bmatrix} \frac{1}{a_{1k}}\Sigma_{1k} & 0 \\ 0 & \frac{1}{a_{2k}}\Sigma_{2k} \end{bmatrix},\ \Sigma_{1k} = Diag(\boldsymbol{\pi}_{1k}) - \boldsymbol{\pi}_{1k}\boldsymbol{\pi}'_{1k}, \text{ and } \Sigma_{2k} = Diag(\boldsymbol{\pi}_{2k}) - \boldsymbol{\pi}_{2k}\boldsymbol{\pi}'_{2k}.$$

Applying the delta method, we obtain $\sqrt{N_+}(\hat{\psi} - \psi) \xrightarrow{d} N(0, V_{\hat{\psi}} = \boldsymbol{\zeta}'\Sigma\boldsymbol{\zeta})$, where $\boldsymbol{\zeta} = (\boldsymbol{\zeta}'_1, \dots, \boldsymbol{\zeta}'_K)'$, $\boldsymbol{\zeta}_k = (\boldsymbol{\zeta}'_{1k}\ \boldsymbol{\zeta}'_{2k})'$, $\boldsymbol{\zeta}_{ik} = (\zeta_{i1k}, \zeta_{i2k}, \dots, \zeta_{icK})'$. We simplify $V_{\hat{\psi}}$ as follows:

$$\begin{aligned} V_{\hat{\psi}} &= \boldsymbol{\zeta}'\Sigma\boldsymbol{\zeta} = \sum_k [\frac{1}{a_{1k}}\boldsymbol{\zeta}'_{1k}\Sigma_{1k}\boldsymbol{\zeta}_{1k} + \frac{1}{a_{2k}}\boldsymbol{\zeta}'_{2k}\Sigma_{2k}\boldsymbol{\zeta}_{2k}] \\ &= \sum_k \frac{1}{a_{1k}}[\sum_{a=1}^{c} \pi_{1ak}\zeta_{1ak}^2 - (\sum_{a=1}^{c} \pi_{1ak}\zeta_{1ak})^2] + \sum_k \frac{1}{a_{2k}}[\sum_{b=1}^{c} \pi_{2bk}\zeta_{2bk}^2 - (\sum_{b=1}^{c} \pi_{2bk}\zeta_{2bk})^2]. \end{aligned}$$

The term $\sum_{a=1}^{c} \pi_{1ak}\zeta_{1ak}$ simplifies to

$$\begin{aligned} &\frac{t_k\psi_k}{\sum_l t_l}[\frac{1}{A_k}\sum_{a=1}^{c}\sum_{b=a+1}^{c}(b-a)\pi_{1ak}\pi_{2bk}] - \frac{\psi}{\sum_l t_l}(a_{1k}^{-1} + a_{2k}^{-1})^{-1}(\sum_{a=1}^{c}\sum_{b=1}^{a-1}(a-b)\pi_{2bk}\pi_{1ak}) \\ = &\frac{t_k\psi_k}{\sum_l t_l}[\frac{1}{A_k}\sum_{a<b}^{c}(b-a)\pi_{1ak}\pi_{2bk}] - \frac{\psi}{\sum_l t_l}(a_{1k}^{-1} + a_{2k}^{-1})^{-1}(\sum_{a>b}^{c}(a-b)\pi_{2bk}\pi_{1ak}) \\ = &\frac{t_k\psi_k}{\sum_l t_l} - \frac{t_k\psi}{\sum_l t_l} = w_k(\psi_k - \psi). \end{aligned}$$

Similarly, the term $\sum_{b=1}^{c} \pi_{2bk}\zeta_{2bk} = w_k(\psi_k - \psi)$.

Finally,

$$
\begin{aligned}
V_{\hat{\psi}} &= \frac{1}{(\sum_l t_l)^2}\sum_k \frac{1}{a_{1k}}\{\sum_{a=1}^{c}\pi_{1ak}[t_k\psi_k\frac{1}{A_k}\sum_{b=a+1}^{c}(b-a)\pi_{2bk} \\
&- \psi(a_{1k}^{-1}+a_{2k}^{-1})^{-1}(\sum_{b=1}^{a-1}(a-b)\pi_{2bk})]^2\} \\
&+ \frac{1}{(\sum_l t_l)^2}\sum_k \frac{1}{a_{2k}}\{\sum_{b=1}^{c}\pi_{2bk}[t_k\psi_k\frac{1}{A_k}\sum_{a=1}^{b-1}(b-a)\pi_{1ak} \\
&- \psi(a_{1k}^{-1}+a_{2k}^{-1})^{-1}(\sum_{a=b+1}^{c}(a-b)\pi_{1ak})]^2\} - \sum_k (a_{1k}^{-1}+a_{2k}^{-1})[w_k(\psi_k-\psi)]^2 .
\end{aligned}
$$

Since $t_k/B_k = (a_{1k}^{-1}+a_{2k}^{-1})^{-1}$, the asymptotic variance can be written as

$$
\begin{aligned}
V_{\hat{\psi}} &= \frac{1}{(\sum_{l=1}^{K} t_l)^2}\{\sum_{k=1}^{K}\frac{t_k^2}{a_{1k}}[\sum_{a=1}^{c}\pi_{1ak}(\frac{\psi_k}{A_k}\sum_{b=a+1}^{c}(b-a)\pi_{2bk} - \frac{\psi}{B_k}\sum_{b=1}^{a-1}(a-b)\pi_{2bk})^2] \\
&\quad + \sum_{k=1}^{K}\frac{t_k^2}{a_{2k}}[\sum_{b=1}^{c}\pi_{2bk}(\frac{\psi_k}{A_k}\sum_{a=1}^{b-1}(b-a)\pi_{1ak} - \frac{\psi}{B_k}\sum_{a=b+1}^{c}(a-b)\pi_{1ak})^2]\} \\
&- \sum_k (a_{1k}^{-1}+a_{2k}^{-1})[w_k(\psi_k-\psi)]^2 ,
\end{aligned}
$$

□

Notice that this expression is valid even if the odds ratios vary among the strata and the collapsings. Under the common cumulative odds ratio assumption, we have $\psi_1 = \psi_2 = \cdots = \psi_K = \psi$. The term $[w_k(\psi_k-\psi)]^2$ in Lemma 2.4.1 is zero and $V_{\hat{\psi}}$ can be written as

$$
\begin{aligned}
V_{\hat{\psi}} &= \frac{\psi^2}{(\sum_{l=1}^{K} t_l)^2}\{\sum_{k=1}^{K} t_k^2[\frac{1}{a_{1k}}\sum_{a=1}^{c}\pi_{1ak}(\frac{1}{A_k}\sum_{b=a+1}^{c}(b-a)\pi_{2bk} - \frac{1}{B_k}\sum_{b=1}^{a-1}(a-b)\pi_{2bk})^2 \\
&\quad + \frac{1}{a_{2k}}\sum_{b=1}^{c}\pi_{2bk}(\frac{1}{A_k}\sum_{a=1}^{b-1}(b-a)\pi_{1ak} - \frac{1}{B_k}\sum_{a=b+1}^{c}(a-b)\pi_{1ak})^2]\} \\
&= \psi^2\frac{\sum_k t_k^2(\frac{1}{a_{1k}}\sum_{a=1}^{c}\pi_{1ak}\eta_{1ak}^2 + \frac{1}{a_{2k}}\sum_{b=1}^{c}\pi_{2bk}\eta_{2bk}^2)}{(\sum_k t_k)^2} = \psi^2\frac{\sum_k t_k^2/\omega_k}{(\sum_k t_k)^2}, \qquad (2.21)
\end{aligned}
$$

where $\omega_k^{-1} = (1/a_{1k})[\sum_{a=1}^{c} \pi_{1ak}\eta_{1ak}^2] + (1/a_{2k})[\sum_{b=1}^{c} \pi_{2bk}\eta_{2bk}^2]$, $\eta_{1ak} = (1/A_k)[\sum_{b=a+1}^{c}(b-a)\pi_{2bk}] - (1/B_k)[\sum_{b=1}^{a-1}(a-b)\pi_{2bk}]$, and $\eta_{2bk} = (1/A_k)[\sum_{a=1}^{b-1}(b-a)\pi_{1ak}] - (1/B_k)[\sum_{a=b+1}^{c}(a-b)\pi_{1ak}]$. The variance expressed in (2.21) can also be obtained using Hauck's procedure.

From Lemma 2.4.1 and under the common cumulative odds ratio assumption, we conclude that

$$\sqrt{N_+}(\hat{\psi} - \psi) \text{ is } AN\left(0, \frac{\psi^2 \sum_k t_k^2/\omega_k}{(\sum_k t_k)^2}\right),$$

where 'AN' denotes an asymptotic normal distribution. We can also express the asymptotic variance as the following:

$$\begin{aligned} N_+ \mathrm{Var}^a(\hat{\psi}) &= \frac{\psi^2 \sum_k t_k^2/\omega_k}{(\sum_k t_k)^2} = V_{\hat{\psi}} \\ &= \frac{\sum_{k=1}^{K} \sum_{j=1}^{c-1} N_+ \mathrm{Var}^a[(R_{jk} - \psi S_{jk})/N_+]}{[\sum_{k=1}^{K} \sum_{j=1}^{c-1} \mathrm{E}^a(S_{jk}/N_+)]^2} \\ &+ \frac{\sum_{k=1}^{K} 2\sum_{j<j'}^{c-1} N_+ \mathrm{Cov}^a[(R_{jk} - \psi S_{jk})/N_+, (R_{j'k} - \psi S_{j'k})/N_+]}{[\sum_{k=1}^{K} \sum_{j=1}^{c-1} \mathrm{E}^a(S_{jk}/N_+)]^2}. \end{aligned} \quad (2.22)$$

An asymptotic variance of $\hat{\psi}$ under Model I follows from the expression (2.21) or (2.22). For Model I, we define an estimator $N_+\hat{\mathrm{V}}\mathrm{ar}(\hat{\psi})$ having the same form as (2.10),

$$N_+ \hat{\mathrm{V}}\mathrm{ar}(\hat{\psi}) = \frac{\sum_{k=1}^{K} \hat{\xi}_k(\hat{\psi})/N_+}{(\sum_{k=1}^{K} \sum_{j=1}^{c-1} S_{jk}/N_+)^2},$$

where $\{\hat{\xi}_k(\hat{\psi})\}$ are defined in the expression (2.10) in Section 2.3. If $N_+\hat{\mathrm{V}}\mathrm{ar}(\hat{\psi})$ is consistent for (2.21) or (2.22), the estimator $\hat{\mathrm{V}}\mathrm{ar}(\hat{\psi})$ is consistent for estimating the variance of $\hat{\psi}$ under both Models I and II.

Next, we illustrate the consistency of our proposed estimator (2.10) for estimating the variance of $\hat{\psi}$ under Model I. From Lemma 2.3.1 and Lemma 2.3.2, we know

that $\mathrm{Var}(R_{jk} - \psi S_{jk}) = Z_{jk} + D_{jk}$ and $\mathrm{Cov}(R_{jk} - \psi S_{jk}, R_{j'k} - \psi S_{j'k}) = Z_{jj'k} + D_{jj'k}$ for all $j < j' \in \{1, \ldots, c-1\}$ and $k \in \{1, \ldots, K\}$. The forms of Z_{jk}, D_{jk}, $Z_{jj'k}$, and $D_{jj'k}$ are given in (2.11), (2.12), (2.17), and (2.18) respectively. Since $\{Z_{jk}/N_+\}$ and $\{Z_{jj'k}/N_+\}$ are $O(1/N_+)$ and $\{D_{jk}/N_+\}$ and $\{D_{jj'k}/N_+\}$ are $O(1)$, D_{jk}/N_+ and $D_{jj'k}/N_+$ are the asymptotically non-negligible portions of the exact $\mathrm{Var}(R_{jk} - \psi S_{jk})/N_+$ and $\mathrm{Cov}(R_{jk} - \psi S_{jk}, R_{j'k} - \psi S_{j'k})/N_+$. Therefore, from the expression of the asymptotic variance in (2.21) and (2.22), we expect that $\psi^2 t_k^2/\omega_k$ equals $\lim_{N_+\to\infty}[(\sum_j D_{jk} + 2\sum_{j<j'} D_{jj'k})/N_+]$. However, the proof that this equality holds will be given in Appendix A.

Since $\{X_{ijk}/n_{ik}\}$ are consistent estimators of $\{\pi_{ijk}\}$ under Model I for $i \in \{1, 2\}$, $j \in \{1, \ldots, c\}$, and $k \in \{1, \ldots, K\}$, it follows by the construction of $\hat{\phi}_{jk}(\psi)$ and $\hat{\phi}_{jj'k}(\psi)$ in the form (2.10) that $\hat{\phi}_{jk}(\psi)/N_+$ and $\hat{\phi}_{jj'k}(\psi)/N_+$ converges to $\lim_{N_+\to\infty} D_{jk}/N_+$ and $\lim_{N_+\to\infty} D_{jj'k}/N_+$, respectively. Thus, $(\sum_j \hat{\phi}_{jk}(\psi) + 2\sum_{j<j'} \hat{\phi}_{jj'k}(\psi))/N_+ = \hat{\xi}_k(\psi)/N_+$ converges to $\lim_{N_+\to\infty}(\sum_j D_{jk} + 2\sum_{j<j'} D_{jj'k})/N_+$. Also, since $\psi^2 t_k^2/\omega_k$ equals $\lim_{N_+\to\infty}[(\sum_j D_{jk} + 2\sum_{j<j'} D_{jj'k})/N_+]$ given in Appendix A, $\sum_{k=1}^K \hat{\xi}_k(\psi)/N_+$ converges to the numerator of (2.21). Furthermore, $\sum_{j=1}^{c-1} S_{jk}/N_+$ converges to t_k under Model I, since

$$\sum_{j=1}^{c-1} S_{jk}/N_+ = \frac{1}{N_+}\sum_j \frac{X^*_{2jk}(n_{1k} - X^*_{1jk})}{N_k} = \frac{n_{1k}n_{2k}}{N_+}\frac{1}{N_k}\sum_j \frac{X^*_{2jk}}{n_{2k}}\left(1 - \frac{X^*_{1jk}}{n_{1k}}\right)$$

$$= \frac{1}{N_+}\left(\frac{1}{n_{1k}} + \frac{1}{n_{2k}}\right)^{-1}\sum_j \frac{X^*_{2jk}}{n_{2k}}\left(1 - \frac{X^*_{1jk}}{n_{1k}}\right) \xrightarrow{p} \left(\frac{1}{a_{1k}} + \frac{1}{a_{2k}}\right)^{-1}\sum_j \pi^*_{2jk}(1 - \pi^*_{1jk}) = t_k\,.$$

Again, since $\hat{\psi} \xrightarrow{p} \psi$, and $\sum_k \hat{\xi}_k(\psi)$ is a smooth function of ψ, we conclude that $N_+\hat{\mathrm{Var}}(\hat{\psi})$ is a consistent estimator of the limiting variance (2.21) of $\sqrt{N_+}\hat{\psi}$. Thus, from Sections 2.3 and 2.4, our proposed estimator $\hat{\mathrm{Var}}(\hat{\psi})$ as given in (2.10) is a consistent estimator for estimating the variance of $\hat{\psi}$ under both Models I and II.

2.5 Ordinal Test of Conditional Independence

We first introduce the ordinary Mantel–Haenszel test of conditional independence for $2\times 2\times K$ contingency tables. Then, we present a generalized Mantel–Haenszel test for $r\times c\times K$ contingency tables. Finally, we propose a test of conditional independence that refers to the common cumulative odds ratio assumption.

For $2\times c\times K$ contingency tables, conditional independence means that there is no association between the explanatory variable and the response variable within each stratum; that is, the response distributions for the two groups of the explanatory variable are the same, given the strata. For the binary response case, under conditional independence, given the row and column marginal totals in stratum k, X_{11k} follows the hypergeometric distribution with $\mathrm{E}(X_{11k}) = n_{1k}m_{1k}/N_k$ and variance $\mathrm{Var}(X_{11k}) = n_{1k}n_{2k}m_{1k}m_{2k}/(N_k^2(N_k-1))$. Since the strata are independent samples, $\sum_k X_{11k}$ has mean $\sum_k \mathrm{E}(X_{11k})$ and variance $\sum_k \mathrm{Var}(X_{11k})$. Mantel and Haenszel (1959) proposed the test statistic

$$M^2 = \frac{\left[\left|\sum_k X_{11k} - \sum_k \mathrm{E}(X_{11k})\right| - 0.5\right]^2}{\sum_k \mathrm{Var}(X_{11k})}$$

for testing conditional independence. Under conditional independence, this statistic has approximately a chi-squared distribution with 1 degree of freedom for both Models I and II. A continuity correction 0.5 was suggested because $\sum_k X_{11k} - \sum_k \mathrm{E}(X_{11k})$ is discrete, with unit-spaced potential values.

Mantel and Haenszel (1959) showed that $\hat{\psi}_{\mathrm{MH}}$ (2.2) equals unity only when $\sum_k X_{11k} = \sum_k \mathrm{E}(X_{11k})$. So, the statistic M^2 without the continuity correction equals zero when the Mantel–Haenszel estimate equals one. The test statistic M^2 can be conducted as a significance test of the departure of the weighted average of the stratum-specific odds

ratio from unity. However, the test is inappropriate when the association changes dramatically across the strata. The test statistic M^2 is also called the *Cochran–Mantel–Haenszel* statistic, since Cochran (1954) proposed a similar statistic by treating rows as independent binomial samples.

The sample size discussions for M^2 to be approximately chi–squared are available in Mantel and Fleiss (1980), Breslow and Day (1980, p. 139), Fleiss (1981), Koch *et al.* (1985), and Agresti (1990, p. 232). Mantel and Fleiss (1980) calculated the maximum and minimum values of $\sum_k X_{11k}$ based on fixed marginal totals in 2×2 contingency tables. The maximum value is $\sum_k \min(n_{1k}, m_{1k})$ and the minimum value is $\sum_k \max(0, m_{1k} - n_{2k})$. Under the null hypothesis, $\sum_k \mathrm{E}(X_{11k})$ has to be at least five units away from both these extremes to ensure an approximately chi–squared distribution. When each stratum has a single matched pair of subjects, strata in which both subjects make the same response do not contribute to M^2. In this case, the total number of pairs of subjects with different responses must be relatively large.

Birch (1965), Landis *et al.* (1978), and Mantel and Byar (1978) generalized the Mantel–Haenszel statistic to $r \times c \times K$ contingency tables. In each stratum, the explanatory variable has r categories and the response variable has c categories. Conditional on the row and column marginal totals in each stratum, $\boldsymbol{X}_k = (X_{11k}, X_{12k}, \ldots, X_{1(c-1)k}, \ldots, X_{(r-1)(c-1)k})'$ has the multiple hypergeometric distribution under the hypothesis of conditional independence. The mean $\mathrm{E}(\boldsymbol{X}_k)$ and variance and covariance matrix $\boldsymbol{C}_k$ with elements $\{\mathrm{Cov}(X_{ijk}, X_{i'j'k})\}$ are as follows:

$$\mathrm{E}(\boldsymbol{X}_k) = (n_{1k}m_{1k}, n_{1k}m_{2k}, \ldots, n_{1k}m_{(c-1)k}, \ldots, n_{(r-1)k}m_{(c-1)k})'/N_k\,,$$

$$\mathrm{Cov}(X_{ijk}, X_{i'j'k}) = \frac{n_{ik}(\delta_{ii'}N_k - n_{i'k})m_{jk}(\delta_{jj'}N_k - m_{j'k})}{N_k^2(N_k - 1)}\,, \qquad (2.23)$$

with $\delta_{ab} = 1$ when $a = b$ and $\delta_{ab} = 0$ otherwise. We let $\boldsymbol{X} = \sum_k \boldsymbol{X}_k$, $\mathrm{E}(\boldsymbol{X}) = \sum_k \mathrm{E}(\boldsymbol{X}_k)$, and $\boldsymbol{C} = \sum_k \boldsymbol{C}_k$. The generalized Mantel–Haenszel statistic is

$$M_{\mathrm{G}}^2 = (\boldsymbol{X} - \mathrm{E}(\boldsymbol{X}))'\boldsymbol{C}^{-1}(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})).$$

This test statistic has asymptotically a chi–squared distribution with $(r-1)(c-1)$ degrees of freedom under conditional independence for both Models I and II.

We now return to the case of $2 \times c \times K$ contingency tables with an ordinal response. For a $2 \times c$ table, when we collapse the response into a binary outcome $(\leq j, > j)$ where $j = 1, \ldots, c-1$, the odds ratios are the same for all $c-1$ collapsed 2×2 tables based on the proportional odds assumption. Therefore, for the cumulative odds ratio, we consider the $c-1$ collapsed 2×2 tables instead of the $2 \times c$ contingency table. For the jth collapsed table in the kth stratum, X^*_{1jk} has the hypergeometric distribution with mean $\mathrm{E}(X^*_{1jk}) = n_{1k}m^*_{jk}/N_k$ and variance $\mathrm{Var}(X^*_{1jk}) = n_{1k}n_{2k}m^*_{jk}(N_k - m^*_{jk})/(N_k^2(N_k-1))$, given the marginal totals (n_{1k}, n_{2k}, m^*_{jk}, and $N_k - m^*_{jk}$) and under conditional independence as in the binary case. If we treat all collapsed tables as independent samples, a test statistic for conditional independence can be conducted based on the ordinary Mantel–Haenszel test statistic M^2. Since our proposed common cumulative odds ratio estimate (2.3) is based on this naive assumption, as for the $2 \times 2 \times K$ case, the common cumulative odds ratio estimate $\hat{\psi}$ (2.3) equals unity only when $\sum_k \sum_j X^*_{1jk} = \sum_k \sum_j \mathrm{E}(X^*_{1jk})$. So, we propose a test statistic similar to M^2, with a numerator $\sum_k \sum_j X^*_{1jk} - \sum_k \sum_j \mathrm{E}(X^*_{1jk})$. The variance of $\sum_k \sum_j X^*_{1jk}$ based on the naive assumption is not appropriate. We next adjust the denominator of M^2 to use a more appropriate variance for $\sum_k \sum_j X^*_{1jk}$.

Since $\{X^*_{1jk}\}$ are dependent for different collapsings within each stratum, we modify the denominator of the test statistic M^2 by considering covariances between different collapsings. However, it is not clear what is meant by conditional covariances of $X^*_{11k}, \ldots, X^*_{1(j-1)k}$, which are defined on different sample spaces. The covariance of two random variables makes sense only if they can be defined on a common sample space. A solution that simplifies this problem is given by McCullagh and Nelder

(1989, p. 275). They suggested using the covariances from the multiple hypergeometric distribution of $\boldsymbol{X}_{1k} = (X_{11k}, \ldots, X_{1(c-1)k})'$, conditioning on the row and column marginal totals in the kth $2 \times c$ contingency table. The covariance matrix with elements $\{\mathrm{Cov}(X_{1jk}, X_{1j'k})\}$ is given in (2.23). Since $\{X^*_{1jk}\}$ are linear functions of $\{X_{1jk}\}$, the variance of $\sum_k \sum_j X^*_{1jk}$ can be obtained by the variance and covariance of $\{X_{1jk}\}$ given in (2.23).

Therefore, a proposed test statistic for testing conditional independence under the common cumulative odds ratio is given by

$$M^2_{\mathrm{C}} = \frac{\left[\sum_{k=1}^{K} \sum_{j=1}^{c-1} X^*_{1jk} - \sum_{k=1}^{K} \sum_{j=1}^{c-1} \mathrm{E}(X^*_{1jk})\right]^2}{\sum_{k=1}^{K} [\mathrm{Var}(\sum_{j=1}^{c-1} X^*_{1jk})]}.$$

This test statistic simplifies to

$$M^2_{\mathrm{C}} = \frac{\left[\sum_k [\sum_{j=1}^{c} (c-j) X_{1jk} - \mathrm{E}(\sum_{j=1}^{c} (c-j) X_{1jk})]\right]^2}{\sum_k \mathrm{Var}(\sum_{j=1}^{c} (c-j) X_{1jk})}. \tag{2.24}$$

The form expressed in (2.24) is a special case of the test statistic proposed by Mantel (1963). Suppose we assign monotone scores $\{u_i\}$ to levels of the explanatory variable and $\{v_j\}$ to levels of the response variable. The Mantel (1963) *score test statistic*, denoted by M^2_{S}, is given as

$$M^2_{\mathrm{S}} = \frac{\left[\sum_k [\sum_{i=1}^{2} \sum_{j=1}^{c} u_i v_j X_{ijk} - \mathrm{E}(\sum_{i=1}^{2} \sum_{j=1}^{c} u_i v_j X_{ijk})]\right]^2}{\sum_k \mathrm{Var}(\sum_{i=1}^{2} \sum_{j=1}^{c} u_i v_j X_{ijk})}.$$

It has approximately a chi-squared null distribution with 1 degree of freedom. It is used to test conditional independence and is sensitive to a monotone conditional relation between the explanatory variable and the response variable, with the same direction at each stratum. The proposed test statistic M^2_{C} (2.24) is a special case of the Mantel score test statistic M^2_{S} with scores $\{u_i = 2 - i, i = 1, 2\}$ and $\{v_j =$

$c - j, j = 1, \ldots, c\}$. Under conditional independence and given the marginal totals, $E(\sum_{i=1}^{2} \sum_{j=1}^{c} u_i v_j X_{ijk}) = (\sum_i u_i n_{ik})(\sum_j v_j m_{jk}) / N_k$ and $\mathrm{Var}(\sum_{i=1}^{2} \sum_{j=1}^{c} u_i v_j X_{ijk}) = [1/(N_k - 1)] \times [\sum_i u_i^2 n_{ik} - (\sum_i u_i n_{ik})^2 / N_k] \times [\sum_j v_j^2 m_{jk} - (\sum_j v_j m_{jk})^2 / N_k]$. Therefore, the test statistic in expression (2.24) is

$$M_C^2 = \frac{\left[\sum_k [\sum_{j=1}^{c} (c-j) X_{1jk} - n_{1k} (\sum_j (c-j) m_{jk}) / N_k]\right]^2}{\sum_k [1/(N_k - 1)][n_{1k} - n_{1k}^2 / N_k][\sum_j (c-j)^2 m_{jk} - (\sum_j (c-j) m_{jk})^2 / N_k]},$$

which also has approximately a chi-squared null distribution with 1 degree of freedom.

2.6 An Example

Table 2.2, from Stuart (1953), describes unaided distance vision for a sample of women aged 30–39. The table presents the left and right eye-testing grades with ordered categories (highest, second, third, and lowest) for 7477 women.

Table 2.2. Unaided distance vision for 7477 women aged 30–39.

	Left Eye Grade				
Right Eye Grade	Highest	Second	Third	Lowest	Total
Highest	1520	266	124	66	1976
Second	234	1512	432	78	2256
Third	117	362	1772	205	2456
Lowest	36	82	179	492	789
Total	1907	2222	2507	841	7477

Table 2.2 can be re-expressed as a $2 \times 4 \times 7477$ contingency table, with a single matched pair forming a stratum. For example, there are 1520 strata (women) with a pattern as Table 2.3. Let W be the control variable (the strata) with level k, Y

Table 2.3. A stratum with both eyes having highest grade.

Factor (Z)	Response (Y) level j				
level i	1	2	3	4	Total
1 (Right Eye Grade)	1	0	0	0	1
2 (Left Eye Grade)	1	0	0	0	1
Total	2	0	0	0	2

denote the response (the grade) with level j, and Z denote the explanatory variable (the eyes) with level i, where $k = 1, \ldots, 7477$, $j = 1$ (highest), 2 (second), 3 (third), 4 (lowest) and $i = 1$ (right eye grade), 2 (left eye grade). We can fit the common proportional odds model (2.1) by the ML method or using the proposed Mantel–Haenszel–type estimator to compare the unaided distance vision distributions of the right and left eyes.

For the data given in Table 2.2, the proposed Mantel–Haenszel–type estimate for the common cumulative odds ratio (2.3) equals $\hat{\psi} = 1.175$, and the estimated variance (2.10) equals 0.0030. Thus, $\hat{\Delta} = \log(\hat{\psi}) = 0.161$ has estimated standard error 0.047 $(= \sqrt{0.0030}/1.175)$, applying the delta method. For testing the null hypothesis of marginal homogeneity, we use the Wald statistic $= (0.161/0.047)^2 = 11.7$, having $p-$value less than 0.001. We conclude that there is evidence that the right and left eyes have different vision distributions. The estimated odds that the grade is below any fixed level is $\exp(0.161) = 1.175$ times higher for the right eye than for the left eye; that is, the women's right eyes tend to have better unaided distance vision than the left eyes.

A software FORTRAN program written by Lang (1992) is available for ML fitting of the common proportional odds model (2.1). Standard software for proportional odds models, such as PROC LOGISTIC in SAS or PLUM (McCullagh, 1980), applies

to a simpler proportional odds model than (2.1),

$$\log(\frac{\pi^*_{ijk}}{1-\pi^*_{ijk}}) = \alpha_j + \Delta I_{\{i=1\}} + \beta_k \ , \tag{2.25}$$

where $i = 1, 2$, $j = 1, 2, 3, 4$, and $k = 1, \ldots, 7477$.

For the data, the unconditional ML estimate for the simpler model (2.25) equals $\hat{\Delta} = 0.211$, with estimated standard error 0.047; it equals $\hat{\Delta} = 0.296$, with estimated standard error 0.061 for the common proportional odds model (2.1). Table 2.4 shows the estimate and estimated standard error for the unconditional ML method and for the proposed Mantel–Haenszel–type method. The unconditional ML estimates are

Table 2.4. Comparison between the unconditional ML estimates and proposed Mantel–Haenszel–type estimate

	$\hat{\Delta}$	s. e.
unconditional ML for Model (2.25)	0.211	0.047
unconditional ML for Model (2.1)	0.296	0.061
Mantel–Haenszel–type	0.161	0.047

about double the Mantel–Haenszel–type estimate. This is not surprising, because when the number of strata is large and the data are sparse the unconditional ML estimate for a $2 \times 2 \times K$ contingency table tends to overestimate the log odds ratio. Also, the unconditional ML estimate in a $2 \times 2 \times K$ contingency table table with one observation in each row of each partial table is double the conditional ML estimate, which is an alternative consistent estimator for that case.

For $2 \times c \times K$ contingency tables in which each stratum consists of a single pair of subjects, the proposed Mantel–Haenszel–type estimator for the cumulative odds ratio (2.3) and the proposed estimated variance (2.10) simplify dramatically. We can express the estimators directly in terms of counts in a $c \times c$ squared table that

expresses the joint ratings of the two observations from the partial tables. Table 2.2 is an example of such a table. Let x_{ij} be the sample count for the ith row and the jth column in such a table. The proposed estimate for $\log(\psi)$ and its estimated asymptotic variance simplify to:

$$\log(\hat{\psi}) = \log\{[\sum_{i<j}(j-i)x_{ij}]/[\sum_{i>j}(i-j)x_{ij}]\},$$

$$\hat{\mathrm{V}}\mathrm{ar}[\log(\hat{\psi})] = \sum_{i<j}(j-i)^2 x_{ij}/[\sum_{i<j}(j-i)x_{ij}]^2 + \sum_{i>j}(i-j)^2 x_{ij}/[\sum_{i>j}(i-j)x_{ij}]^2.$$

The simplifications are shown in Appendix B. Agresti and Lang (1993) proposed this simplified estimator and the simplified estimated variance for a proportional odds model for subject-specific distributions.

CHAPTER 3
EVALUATION OF ORDINAL ODDS RATIO ESTIMATORS

3.1 Introduction

The proposed Mantel–Haenszel–type estimator (2.3) for $2 \times c \times K$ contingency tables is the ordinary Mantel–Haenszel estimator for $2 \times 2 \times (c-1)K$ contingency tables obtained by naively treating different collapsed tables for each stratum as independent samples. However, the standard error for the estimator based on the naive assumption is not appropriate. In Chapter 2, we proposed an estimator of the variance (2.10) based on the independent multinomial sampling assumption. Our common cumulative odds ratio estimator (2.3) and its variance estimator (2.10) are consistent under both the large–stratum and the sparse–data cases, but our estimator may have a loss of efficiency compared to the ML estimator under the large–stratum limiting model, because the proposed estimator is based on the naive assumption.

Under the large–stratum situation with a fixed number of strata, ML estimators have good asymptotic properties (Lehmann, 1983, Chapter 6). The unconditional ML estimator for the common cumulative odds ratio based on the common proportional odds model (2.1) is asymptotically efficient under Model I. In this chapter, we will study the loss of efficiency for the proposed estimator compared to the unconditional ML estimator under the large–stratum situation. We use the Pitman asymptotic relative efficiency (ARE) to measure local relative efficiency for the two estimators. The results show that our estimator has little loss of efficiency.

We then use a simulation study to show that the proposed estimator is much better than the unconditional ML estimator under the sparse–data situation. The

simulation study generates observations from the simpler common proportional odds model (2.25). A reason for using the simpler common proportional odds model (2.25) instead of the general common proportional odds model (2.1) is because software for ML estimation is widely available for the simpler model. The study provides a variety of scenarios, ranging from ones for which large-strata asymptotics should work well (*e. g.*, $K = 5$ and $n_{ik} = 20$) to ones for which sparse-data limiting asymptotics would seem more appropriate (*e. g.*, $K = 25$ and $n_{ik} \leq 3$). From these efficiency and simulation studies, we conclude that our estimator is nearly as good as the unconditional ML estimator under Model I, but it is much better than the unconditional ML estimator under Model II.

In Chapter 2, the procedures for obtaining a dually consistent variance estimator apply under the hypothesis that the cumulative odds ratio is constant across the strata. This assumption may not hold in general. This chapter reviews several possible tests of a homogeneous odds ratio for $2 \times 2 \times K$ cases. Generalizing these tests, we propose several similar tests of a homogeneous cumulative odds ratio. Our proposed estimator also assumes that the odds ratios are constant for the different possible collapsings within each stratum. In this chapter, we also propose a test of homogeneity across the collapsings. This test is appropriate under both limiting models.

If the assumption of the common cumulative odds ratio does not hold, the common cumulative odds ratio estimator (2.3) provides a summary of the strata–specific cumulative odds ratios, but the estimated variance (2.10) for that estimator will be inappropriate. This chapter presents a valid estimated variance only under the large–stratum situation.

In Section 3.2, we analyze the Pitman ARE for the proposed estimator, compared to the unconditional ML estimator under Model I. In Section 3.3, we study the performance of the proposed estimator by a simulation study. Section 3.4 discusses several tests of the assumption of common cumulative odds ratio. Finally, in Section 3.5, we

introduce a large–stratum estimated variance for estimator (2.3) that applies when the common cumulative odds ratio assumption does not hold.

3.2 Efficiency

The unconditional ML estimator for the common odds ratio based on the proportional odds model (2.1) is asymptotically efficient under Model I. Since the proposed estimator (2.3) naively treats the collapsings as independent samples, it may have a loss of efficiency. We now study the asymptotic relative efficiency for our estimator compared to the unconditional ML estimator, under the large–stratum situation. The relative efficiency can be interpreted in terms of the sample sizes required for two types of estimators to estimate a parameter with comparable accuracy. Because the unconditional ML estimator for the common proportional odds model (2.1) does not have a closed form, instead of considering the ML estimator, we consider an asymptotically equivalent test statistic for the null hypothesis. Then, we compare our test statistic to the ML equivalent test statistic. The criterion used is Pitman ARE.

Consider a test of conditional independence; that is, $\psi = 1$ or $\log(\psi) = \Delta = 0$. A test statistic based on the proposed estimator (2.3) and its asymptotic distribution can be easily established under the large–stratum situation. It is asymptotically normally distributed, applying the delta method shown in Section 2.4. Unfortunately, the ML estimator and its asymptotic variance for the common proportional odds model (2.1) do not have closed forms. However, when there is a single stratum, the efficient score test for the null hypothesis based on model (2.1) is equivalent to the Wilcoxon average rank test (McCullagh, 1980; McCullagh and Nelder, 1989, p. 188). The efficient score test, due to R. A. Fisher and C. R. Rao, is based on the log–likelihood derivative with respect to Δ evaluated under the null hypothesis. For the K strata case, we can construct an equivalent efficient score test based on model (2.1) that generalizes the Wilcoxon average rank test. First, we study the Pitman ARE for

the proposed estimator compared to the ML estimator based on a single stratum. Secondly, we construct an equivalent efficient score test and study the Pitman ARE for the K strata case.

The Wilcoxon average rank statistic is equivalent to a rescaled Kendall's tau statistic for discrete data, called *Kendall's tau–a*, which measures the difference between the proportions of concordant and discordant pairs (Lehmann, 1975, p. 22). For calculation simplicity, we use Kendall's tau–a instead of the Wilcoxon average rank statistic. Here, we consider a *rescaled* Kendall's tau–a, which is the difference between the numbers of concordant and discordant pairs. Indeed, Lehmann showed that the Wilcoxon average rank statistic equals the rescaled Kendall's tau–a plus a constant, when the row marginal totals, n_1 and n_2, are fixed. For the one stratum case, a test of the hypothesis has no difference based on Kendall's tau–a or the rescaled Kendall's tau–a. However, for the K strata case, an equivalent efficient score test discussed later generalizes the rescaled Kendall's tau–a. For this reason, we now consider the rescaled Kendall's tau–a.

For comparing the local asymptotic relative efficiency of our proposed estimator and the ML estimator under the one stratum case, two tests are considered, based on the following statistics:

(1) the proposed estimator:

$$\begin{aligned}
\hat{\Delta} = \log(\hat{\psi}) &= \log(\sum_{j=1}^{c-1} X_{1j}^{*}(n_2 - X_{2j}^{*})) - \log(\sum_{j=1}^{c-1} X_{2j}^{*}(n_1 - X_{1j}^{*})) \\
&= \log(\sum_{j=1}^{c-1} p_{1j}^{*}(1 - p_{2j}^{*})) - \log(\sum_{j=1}^{c-1} p_{2j}^{*}(1 - p_{1j}^{*})) \\
&= \log(\sum_{a<b=1}^{c} (b-a)p_{1a}p_{2b}) - \log(\sum_{a>b=1}^{c} (a-b)p_{1a}p_{2b})
\end{aligned}$$

(2) rescaled Kendall's tau-a:

$$\hat{\tau} = \sum_{a=1}^{c-1} X_{1a}\Big(\sum_{b=a+1}^{c} X_{2b}\Big) - \sum_{b=1}^{c-1} X_{2b}\Big(\sum_{a=b+1}^{c} X_{1a}\Big)$$

$$= (n_1 n_2)\Big[\sum_{a=1}^{c-1} p_{1a}\Big(\sum_{b=a+1}^{c} p_{2b}\Big) - \sum_{b=1}^{c-1} p_{2b}\Big(\sum_{a=b+1}^{c} p_{1a}\Big)\Big].$$

Notice that Kendall's tau-a is defined as $2\hat{\tau}/(N(N-1))$, where $N = n_1 + n_2$.

For the one stratum case, the proportional odds model considered is

$$\log\Big(\frac{\pi^*_{ij}}{1-\pi^*_{ij}}\Big) = \alpha_j + \Delta I_{\{i=1\}}\ ,\ i = 1,2,\ j = 1,\ldots,c-1\,. \tag{3.1}$$

Based on the model, the cell probabilities $\{\pi_{ij}\}$ have the following structure:

$$\pi_{ij} = \frac{\exp(\alpha_j + \Delta I_{\{i=1\}})}{1+\exp(\alpha_j + \Delta I_{\{i=1\}})} - \frac{\exp(\alpha_{j-1} + \Delta I_{\{i=1\}})}{1+\exp(\alpha_{j-1} + \Delta I_{\{i=1\}})}\ , \tag{3.2}$$

where $i = 1,2$, $j = 1,\ldots,c$, and $-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_{c-1} < \alpha_c = \infty$. Later, we will use these probabilities to obtain the Pitman ARE.

Let ARE(S, T) denote the Pitman asymptotic relative efficiency for a test statistic S compared to a test statistic T. The ARE(S, T) is the limiting ratio of sample sizes required to achieve the same limiting power against the same sequence of alternatives (converging to the null) when the limiting significance levels of the two tests are equal. The limiting ratio is the limit of the sample size required for T divided by the sample size required for S. When ARE(S, T) is less than 1, the sample size required for S is larger than the sample size required for T to achieve the same limiting power. So, the test statistic T is more efficient than the test statistic S. The interpretation of ARE(S, T) is simple. In order to evaluate ARE(S, T), we can use Noether's (1955) theorem.

Noether's theorem is summarized by Puri and Sen (1971, section 3.8) and Randles and Wolfe (1991, section 5.2). Simon (1978) used this theorem to find the efficacies

of several measures of association for ordinal contingency tables, including Kendall's tau–a. The efficacy of Kendall's tau–a for our case is different from Simon's results because the cell probabilities are based on the proportional odds model, whereas Simon considered a general ϵ–contamination model $\pi_{ij} = (1-\epsilon)\alpha_i\beta_j + \epsilon\gamma_{ij}$. However, the arguments are similar.

Suppose that there is a null hypothesis $H(0)$, and let an alternative hypothesis $H(\Delta)$ approach $H(0)$ as the sample size n grows. We summarize Noether's theorem as follows. Consider two test statistics S_n and T_n. If (1) $(S_n - \mu_{S_n}(\Delta))/\sigma_{S_n}(\Delta)$ and $(T_n - \mu_{T_n}(\Delta))/\sigma_{T_n}(\Delta)$ have the same limiting distributions for all Δ, (2) $\lim_{\Delta\to 0}[\sigma_{S_n}(\Delta)/\sigma_{S_n}(0)] = \lim_{\Delta\to 0}[\sigma_{T_n}(\Delta)/\sigma_{T_n}(0)] = 1$, (3) $\mu_{S_n}(\Delta)$ and $\mu_{T_n}(\Delta)$ are differentiable with $\mu'(0)$ being μ' evaluated at $\Delta = 0$, and (4) $\lim_{\Delta\to 0}[\mu'_{S_n}(\Delta)/\mu'_{S_n}(0)] = \lim_{\Delta\to 0}[\mu'_{T_n}(\Delta)/\mu'_{T_n}(0)] = 1$, then the efficacy of S is defined as

$$\mathrm{Eff}(S) = \lim_{n\to\infty} \frac{\mu'_{S_n}(0)}{\sqrt{n\sigma^2_{S_n}(0)}},$$

and the Pitman ARE for S compared to T is $\mathrm{ARE}(S, T) = (\mathrm{Eff}(S)/\mathrm{Eff}(T))^2$.

The proportional odds model (3.1) discussed here is a multi–parameter model. The parameters are $\Delta, \alpha_1, \ldots, \alpha_{c-1}$. Let $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_{c-1}, \Delta)'$, and $\boldsymbol{\theta}_0 = (\alpha_1, \ldots, \alpha_{c-1}, 0)'$. We let an alternative hypothesis $H(\boldsymbol{\theta})$ approach a null hypothesis $H(\boldsymbol{\theta}_0)$ as the sample size grows. Notice that this limit keeps $\alpha_1, \ldots, \alpha_{c-1}$ fixed and lets $\Delta \to 0$. Noether's theorem can be proved by a Taylor series expansion of $\mu(\boldsymbol{\theta})$ around $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. The Taylor series expansion yields

$$\mu(\boldsymbol{\theta}) = \mu(\boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\boldsymbol{\mu}'(\boldsymbol{\theta}^*), \quad \boldsymbol{\theta}_0 < \boldsymbol{\theta}^* < \boldsymbol{\theta}, \tag{3.3}$$

where $\boldsymbol{\mu}' = (\partial\mu/\partial\alpha_1, \partial\mu/\partial\alpha_2, \ldots, \partial\mu/\partial\Delta)'$. The Taylor series expansion (3.3) simplifies to

$$\mu(\boldsymbol{\theta}) = \mu(\boldsymbol{\theta}_0) + (\Delta - 0)\mu'(\boldsymbol{\theta}^*),$$

where $\mu'(\boldsymbol{\theta}^*)$ is $\partial\mu/\partial\Delta$ evaluated at $\boldsymbol{\theta}^*$. The details of proof are similar to the single parameter case, given by Randles and Wolfe (1991, p. 148). Therefore,

Noether's theorem for a multi–parameter model is modified as follows: If (1) $(S_n - \mu_{S_n}(\boldsymbol{\theta}))/\sigma_{S_n}(\boldsymbol{\theta})$ and $(T_n - \mu_{T_n}(\boldsymbol{\theta}))/\sigma_{T_n}(\boldsymbol{\theta})$ have the same limiting distributions for all $\boldsymbol{\theta}$, (2) $\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0}[\sigma_{S_n}(\boldsymbol{\theta})/\sigma_{S_n}(\boldsymbol{\theta}_0)] = \lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0}[\sigma_{T_n}(\boldsymbol{\theta})/\sigma_{T_n}(\boldsymbol{\theta}_0)] = 1$, (3) $\mu_{S_n}(\boldsymbol{\theta})$ and $\mu_{T_n}(\boldsymbol{\theta})$ are differentiable with $\mu'(\boldsymbol{\theta}_0)$ being μ' evaluated at $\boldsymbol{\theta}_0$, and (4) $\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0}[\mu'_{S_n}(\boldsymbol{\theta})/\mu'_{S_n}(\boldsymbol{\theta}_0)] = \lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0}[\mu'_{T_n}(\boldsymbol{\theta})/\mu'_{T_n}(\boldsymbol{\theta}_0)] = 1$, then the efficacy of S is defined as

$$\mathrm{Eff}(S) = \lim_{n\to\infty} \frac{\mu'_{S_n}(\boldsymbol{\theta}_0)}{\sqrt{n\sigma^2_{S_n}(\boldsymbol{\theta}_0)}},$$

and the Pitman ARE for S compared to T is $\mathrm{ARE}(S, T) = (\mathrm{Eff}(S)/\mathrm{Eff}(T))^2$.

Under the large–stratum situation, sample proportions are asymptotically normally distributed by the Central Limit Theory. Using the delta method, it is easy to derive asymptotic distributions of both the proposed estimator and the rescaled Kendall's tau–a. They are both asymptotically normal, and all conditions in Noether's theorem hold. From Section 2.4, we know that for the one stratum case, $\sqrt{N}(\hat{\Delta} - \Delta)$ is $AN(0, \omega^{-1})$, where $\omega^{-1} = (1/a_1)[\sum_{a=1}^{c} \pi_{1a}\eta_{1a}^2] + (1/a_2)[\sum_{b=1}^{c} \pi_{2b}\eta_{2b}^2]$, $\eta_{1a} = (1/A)[\sum_{b=a+1}^{c} (b-a)\pi_{2b}] - (1/B)[\sum_{b=1}^{a-1}(a-b)\pi_{2b}]$, $\eta_{2b} = (1/A)[\sum_{a=1}^{b-1}(b-a)\pi_{1a}] - (1/B)[\sum_{a=b+1}^{c}(a-b)\pi_{1a}]$, $A = \sum_{a<b}(b-a)\pi_{1a}\pi_{2b}$, $B = \sum_{a>b}(a-b)\pi_{1a}\pi_{2b}$, $a \in \{1, \ldots, c\}$, $b \in \{1, \ldots, c\}$, $N = n_1 + n_2$, and $Na_1 = n_1$, $Na_2 = n_2$. Applying Noether's theorem, the numerator of the efficacy of $\hat{\Delta}$ is $\mu'_{\hat{\Delta}}(\boldsymbol{\theta}_0) = 1$. Let $\{\pi^0_{1a}\}$ be the cell probabilities under the null hypothesis. Under the null hypothesis, we have $\pi^0_{1a} = \pi^0_{2a}$ for $a \in \{1, \ldots, c\}$. Therefore, $A = B = A^0 = \sum_{j=1}^{c-1} \pi^{0*}_{1j}(1 - \pi^{0*}_{1j})$ and $\sum_{a=1}^{c} \pi^0_{1a}(\eta^0_{1a})^2 = \sum_{b=1}^{c} \pi^0_{2b}(\eta^0_{2b})^2 = (1/A^0)^2 \sum_{a=1}^{c} \pi^0_{1a}(\sum_{b=a+1}^{c}(b-a)\pi^0_{1b} - \sum_{b=1}^{a-1}(a-b)\pi^0_{1b})^2$ under the null hypothesis. Thus, the denominator $N\sigma^2_{\hat{\Delta}}(\boldsymbol{\theta}_0)$

$$= \left(\frac{1}{a_1} + \frac{1}{a_2}\right)\frac{1}{(A^0)^2}\left[\sum_{a=1}^{c} \pi^0_{1a}\Big(\sum_{b=a+1}^{c}(b-a)\pi^0_{1b} - \sum_{b=1}^{a-1}(a-b)\pi^0_{1b}\Big)^2\right].$$

Then, the efficacy of $\hat{\Delta}$ is

$$\mathrm{Eff}(\hat{\Delta}) = \frac{A^0}{(\frac{1}{a_1}+\frac{1}{a_2})^{\frac{1}{2}}[\sum_{a=1}^{c}\pi_{1a}^0(\sum_{b=a+1}^{c}(b-a)\pi_{1b}^0 - \sum_{b=1}^{a-1}(a-b)\pi_{1b}^0)^2]^{\frac{1}{2}}}.$$

Next, we derive the efficacy of the rescaled Kendall's tau-a, $\hat{\tau}$. Because of properties of the asymptotic distribution of $(T_n - \mu_{T_n}(\boldsymbol{\theta}))/\sigma_{T_n}(\boldsymbol{\theta})$, the scale $n_1 n_2$ is not relevant. So, we consider $\hat{\tau} = \sum_{a=1}^{c-1} p_{1a}(\sum_{b=a+1}^{c} p_{2b}) - \sum_{b=1}^{c-1} p_{2b}(\sum_{a=b+1}^{c} p_{1a})$. The mean $\mu_{\hat{\tau}}$ is $[\sum_{a=1}^{c-1}\pi_{1a}(\sum_{b=a+1}^{c}\pi_{2b}) - \sum_{b=1}^{c-1}\pi_{2b}(\sum_{a=b+1}^{c}\pi_{1a})]$. Using the probabilities (3.2) under the proportional odds model, the derivative of $\mu_{\hat{\tau}}$ with respect to Δ is

$$\begin{aligned}\mu'_{\hat{\tau}}(\theta) &= \sum_{a=1}^{c-1}\Big(\frac{\exp(\alpha_a+\Delta)}{[1+\exp(\alpha_a+\Delta)]^2} - \frac{\exp(\alpha_{a-1}+\Delta)}{[1+\exp(\alpha_{a-1}+\Delta)]^2}\Big)\Big(\sum_{b=a+1}^{c}\pi_{2b}\Big) \\ &- \sum_{b=1}^{c-1}\pi_{2b}\Big[\sum_{a=b+1}^{c}\Big(\frac{\exp(\alpha_a+\Delta)}{[1+\exp(\alpha_a+\Delta)]^2} - \frac{\exp(\alpha_{a-1}+\Delta)}{[1+\exp(\alpha_{a-1}+\Delta)]^2}\Big)\Big],\end{aligned}$$

where $\exp(\alpha_c+\Delta)/[1+\exp(\alpha_c+\Delta)]^2 = 0$. Under the null hypothesis, we can simplify $\mu'_{\hat{\tau}}(\boldsymbol{\theta}_0)$ to

$$\begin{aligned}&\mu'_{\hat{\tau}}(\boldsymbol{\theta}_0) \\ &= \pi_{11}^0(1-\pi_{11}^0)\sum_{b=2}^{c}\pi_{1b}^0 + \sum_{a=2}^{c-1}\pi_{1a}^{0*}(1-\pi_{1a}^{0*})\sum_{b=a+1}^{c}\pi_{1b}^0 - \sum_{a=2}^{c-1}\pi_{1(a-1)}^{0*}(1-\pi_{1(a-1)}^{0*})\sum_{b=a+1}^{c}\pi_{1b}^0 \\ &- \sum_{b=1}^{c-2}\pi_{1b}^0\sum_{a=b+1}^{c-1}\pi_{1a}^{0*}(1-\pi_{1a}^{0*}) + \sum_{b=1}^{c-1}\pi_{1b}^0\sum_{a=b+1}^{c}\pi_{1(a-1)}^{0*}(1-\pi_{1(a-1)}^{0*}) \\ &= \pi_{11}^0(1-\pi_{11}^0)\sum_{b=2}^{c}\pi_{1b}^0 + \sum_{a=2}^{c-1}\pi_{1a}^{0*}(1-\pi_{1a}^{0*})\Big(\sum_{b=a+2}^{c}\pi_{1b}^0 + \pi_{1(a+1)}^0\Big) \\ &- \sum_{a=1}^{c-2}\pi_{1a}^{0*}(1-\pi_{1a}^{0*})\sum_{b=a+2}^{c}\pi_{1b}^0 - \sum_{b=1}^{c-2}\pi_{1b}^0\sum_{a=b+1}^{c-1}\pi_{1a}^{0*}(1-\pi_{1a}^{0*}) + \sum_{b=1}^{c-1}\pi_{1b}^0\sum_{a=b}^{c-1}\pi_{1a}^{0*}(1-\pi_{1a}^{0*}).\end{aligned}$$

Since

$$\sum_{b=1}^{c-1}\pi_{1b}^{0}\sum_{a=b}^{c-1}\pi_{1a}^{0*}(1-\pi_{1a}^{0*}) = \sum_{b=1}^{c-2}\pi_{1b}^{0}\sum_{a=b}^{c-1}\pi_{1a}^{0*}(1-\pi_{1a}^{0*}) + \pi_{1(c-1)}^{0}\pi_{1(c-1)}^{0*}(1-\pi_{1(c-1)}^{0*})$$

$$= \sum_{b=1}^{c-2}\pi_{1b}^{0}\sum_{a=b+1}^{c-1}\pi_{1a}^{0*}(1-\pi_{1a}^{0*}) + \sum_{b=1}^{c-2}\pi_{1b}^{0}\pi_{1b}^{0*}(1-\pi_{1b}^{0*}) + \pi_{1(c-1)}^{0}\pi_{1(c-1)}^{0*}(1-\pi_{1(c-1)}^{0*})$$

$$= \sum_{b=1}^{c-2}\pi_{1b}^{0}\sum_{a=b+1}^{c-1}\pi_{1a}^{0*}(1-\pi_{1a}^{0*}) + \sum_{b=1}^{c-1}\pi_{1b}^{0}\pi_{1b}^{0*}(1-\pi_{1b}^{0*}),$$

the term $\mu_{\tau}'(\boldsymbol{\theta}_0)$ equals

$$\mu_{\tau}'(\boldsymbol{\theta}_0) = \pi_{11}^{0}(1-\pi_{11}^{0})\sum_{b=2}^{c}\pi_{1b}^{0} + \sum_{a=2}^{c-1}\pi_{1a}^{0*}(1-\pi_{1a}^{0*})\pi_{1(a+1)}^{0} - \pi_{11}^{0}(1-\pi_{11}^{0})\sum_{b=3}^{c}\pi_{1b}^{0}$$

$$+\sum_{b=1}^{c-1}\pi_{1b}^{0}\pi_{1b}^{0*}(1-\pi_{1b}^{0*})$$

$$= \pi_{11}^{0}(1-\pi_{11}^{0})\pi_{12}^{0} + \sum_{a=2}^{c-1}\pi_{1a}^{0*}(1-\pi_{1a}^{0*})\pi_{1(a+1)}^{0} + \sum_{b=1}^{c-1}\pi_{1b}^{0}\pi_{1b}^{0*}(1-\pi_{1b}^{0*})$$

$$= \pi_{11}^{0}(1-\pi_{11}^{0})(\pi_{11}^{0}+\pi_{12}^{0}) + \sum_{a=2}^{c-1}\pi_{1a}^{0*}(1-\pi_{1a}^{0*})(\pi_{1(a+1)}^{0}+\pi_{1a}^{0}).$$

Applying the delta method, we can derive the asymptotic variance, $N\sigma_{\hat{\tau}}^{2}(\boldsymbol{\theta})$. Since

$$\eta_{1a} = \frac{\partial\tau}{\partial\pi_{1a}} = \sum_{b=a+1}^{c}\pi_{2b} - \sum_{b=1}^{a-1}\pi_{2b}, \text{ and } \eta_{2b} = \frac{\partial\tau}{\partial\pi_{2b}} = \sum_{a=1}^{b-1}\pi_{1a} - \sum_{a=b+1}^{c}\pi_{1a},$$

we have

$$N\sigma_{\hat{\tau}}^{2}(\boldsymbol{\theta}) = \frac{1}{a_1}[\sum_{a=1}^{c}\pi_{1a}\eta_{1a}^{2} - (\sum_{a=1}^{c}\pi_{1a}\eta_{1a})^{2}] + \frac{1}{a_2}[\sum_{b=1}^{c}\pi_{2b}\eta_{2b}^{2} - (\sum_{b=1}^{c}\pi_{2b}\eta_{2b})^{2}].$$

Under the null hypothesis, $\sum_{a=1}^{c}\pi_{1a}^{0}\eta_{1a}^{0} = \sum_{a=1}^{c}\pi_{1a}^{0}(\sum_{b=a+1}^{c}\pi_{1b}^{0} - \sum_{b=1}^{a-1}\pi_{1b}^{0}) = \sum_{a=1}^{c}\pi_{1a}^{0}$ $\sum_{b=a+1}^{c}\pi_{1b}^{0} - \sum_{b=1}^{c}\pi_{1b}^{0}\sum_{a=b+1}^{c}\pi_{1a}^{0} = 0 = \sum_{b=1}^{c}\pi_{2b}^{0}\eta_{2b}^{0}$, and $\pi_{1a}^{0}(\eta_{1a}^{0})^{2} = \pi_{1a}^{0}(\sum_{b=a+1}^{c}\pi_{1b}^{0} -$

$\sum_{b=1}^{a-1} \pi_{1b}^0)^2 = \pi_{2b}^0(\eta_{2b}^0)^2$. Thus,

$$N\sigma_{\hat{\tau}}^2(\boldsymbol{\theta}_0) = (\frac{1}{a_1} + \frac{1}{a_2})[\pi_{11}^0(\sum_{b=2}^{c} \pi_{1b}^0)^2 + \sum_{a=2}^{c-1} \pi_{1a}^0(\sum_{b=a+1}^{c} \pi_{1b}^0 - \sum_{b=1}^{a-1} \pi_{1b}^0)^2 + \pi_{1c}^0(\sum_{b=1}^{c-1} \pi_{1b}^0)^2].$$

Therefore,

$$\mathrm{Eff}(\hat{\tau}) = \frac{\pi_{11}^0(1-\pi_{11}^0)(\pi_{11}^0+\pi_{12}^0) + \sum_{a=2}^{c-1} \pi_{1a}^{0*}(1-\pi_{1a}^{0*})(\pi_{1(a+1)}^0 + \pi_{1a}^0)}{(\frac{1}{a_1}+\frac{1}{a_2})^{\frac{1}{2}}[\pi_{11}^0(\sum_{b=2}^{c} \pi_{1b}^0)^2 + \sum_{a=2}^{c-1} \pi_{1a}^0(\sum_{b=a+1}^{c} \pi_{1b}^0 - \sum_{b=1}^{a-1} \pi_{1b}^0)^2 + \pi_{1c}^0(\sum_{b=1}^{c-1} \pi_{1b}^0)^2]^{\frac{1}{2}}}.$$

The Pitman ARE for the proposed Mantel–Haenszel–type estimator (MH) compared to the ML estimator is

$$\mathrm{ARE(MH,\ ML)} = [\mathrm{Eff}(\hat{\Delta})/\mathrm{Eff}(\hat{\tau})]^2 = \frac{k_1}{d_1}/\frac{k_2}{d_2}, \tag{3.4}$$

where

$$k_1 = [\sum_{j=1}^{c-1} \pi_{1j}^{0*}(1-\pi_{1j}^{0*})]^2,$$

$$d_1 = \pi_{11}^0(\sum_{b=2}^{c}(b-1)\pi_{1b}^0)^2 + \sum_{a=2}^{c-1} \pi_{1a}^0(\sum_{b=a+1}^{c}(b-a)\pi_{1b}^0 - \sum_{b=1}^{a-1}(a-b)\pi_{1b}^0)^2$$
$$+\pi_{1c}^0(\sum_{b=1}^{c-1}(c-b)\pi_{1b}^0)^2,$$

$$k_2 = [\pi_{11}^0(1-\pi_{11}^0)(\pi_{11}^0+\pi_{12}^0) + \sum_{a=2}^{c-1} \pi_{1a}^{0*}(1-\pi_{1a}^{0*})(\pi_{1(a+1)}^0+\pi_{1a}^0)]^2,$$

$$d_2 = \pi_{11}^0(\sum_{b=2}^{c} \pi_{1b}^0)^2 + \sum_{a=2}^{c-1} \pi_{1a}^0(\sum_{b=a+1}^{c} \pi_{1b}^0 - \sum_{b=1}^{a-1} \pi_{1b}^0)^2 + \pi_{1c}^0(\sum_{b=1}^{c-1} \pi_{1b}^0)^2.$$

We now study the behavior of the ARE for various cases. We first show that when the probabilities alternate; that is, $\pi_{11}^0 = \pi_{1(2j-1)}^0$ for all $j \in \{1,\ldots,[(c+1)/2]\}$ and $\pi_{12}^0 = \pi_{1(2j)}^0$ for all $j \in \{1,\ldots,[c/2]\}$, the ARE(MH, ML)=1, where the notation $[z]$ represents the largest integer which is not larger than z. Let $x = \pi_{11}^0$ and $y = \pi_{12}^0$.

When the number of the response categories is an odd number, we have $(c+1)x/2 + (c-1)y/2 = 1$. Then, the terms d_1, d_2, and k_2 simplify to

$$d_1 = 2\sum_{j=1}^{(c-1)/2} \pi_{1j}^{0}(\frac{c+1}{2} - j)^2,$$

$$d_2 = [2\sum_{j=1}^{(c-1)/2} \pi_{1j}^{0}(\frac{c+1}{2} - j)^2](x+y)^2 = (x+y)^2 d_1,$$

$$k_2 = (x+y)^2\{\sum_{a=1}^{c-1} \pi_{1a}^{0*}(1-\pi_{1a}^{0*})\}^2 = (x+y)^2 k_1.$$

Then, ARE(MH, ML)$= [k_1/d_1]/[((x+y)^2 k_1)/((x+y)^2 d_1)] = 1$. When the number of the response categories is an even number, we have $c(x+y)/2 = 1$. The terms d_1, d_2, and k_2 simplify to

$$d_1 = 2\sum_{j=1}^{(c/2)-1} \pi_{1j}^{0}(\frac{c}{2} - j + \frac{c}{2}y)^2 + \pi_{1(c/2)}^{0}(\frac{c}{2}y)^2,$$

$$d_2 = [2\sum_{j=1}^{(c/2)-1} \pi_{1j}^{0}(\frac{c}{2} - j + \frac{c}{2}y)^2 + \pi_{1(c/2)}^{0}(\frac{c}{2}y)^2](x+y)^2,$$

$$k_2 = (x+y)^2\{\sum_{a=1}^{c-1} \pi_{1a}^{0*}(1-\pi_{1a}^{0*})\}^2 = (x+y)^2 k_1.$$

Thus, ARE(MH, ML)$= [k_1/d_1]/[((x+y)^2 k_1)/((x+y)^2 d_1)] = 1$. So, under the alternate or equal probability ($x = y$) cases, there is no loss of efficiency for the proposed estimator compared to the ML estimator.

Viewing the log likelihood function for the proportional odds model (3.1), we now present intuition for why there is no loss of efficiency under the alternate probability case. The likelihood from a single multinomial observation $(X_1, X_2, \ldots, X_c)$ can be written as the product of $c-1$ quantities

$$\left\{\left(\frac{\pi_1^*}{\pi_2^*}\right)^{X_1^*}\left(\frac{\pi_2^* - \pi_1^*}{\pi_2^*}\right)^{X_2^* - X_1^*}\right\} \times \left\{\left(\frac{\pi_2^*}{\pi_3^*}\right)^{X_2^*}\left(\frac{\pi_3^* - \pi_2^*}{\pi_3^*}\right)^{X_3^* - X_2^*}\right\} \times \ldots$$

$$\times \left\{ \left(\frac{\pi_{c-1}^*}{\pi_c^*} \right)^{X_{c-1}^*} \left(\frac{\pi_c^* - \pi_{c-1}^*}{\pi_c^*} \right)^{X_c^* - X_{c-1}^*} \right\}, \tag{3.5}$$

given by McCullagh (1980). These factors are respectively the probability given X_2^* that the first two cells divide in the ratio $X_1^* : X_2^* - X_1^*$; the probability given X_3^* that the proportion in cell 3 relative to cells 1 and 2 combined is $X_3^* - X_2^* : X_2^*$ and so on for the other components. Again, the likelihood function (3.5) can be written as

$$\left(\frac{\pi_1^*}{\pi_2^* - \pi_1^*} \right)^{X_1^*} \left(\frac{\pi_2^* - \pi_1^*}{\pi_3^* - \pi_2^*} \right)^{X_2^*} \left[\prod_{j=3}^{c-1} \left(\frac{\pi_j^* - \pi_{j-1}^*}{\pi_{j+1}^* - \pi_j^*} \right)^{X_j^*} \right] \left(\frac{1 - \pi_{c-1}^*}{1} \right)^n,$$

where $n = \sum_{j=1}^{c} X_j$. When the probabilities alternate, the log likelihood function is

$$X_1^* \log(x/y) - X_2^* \log(x/y) + X_3^* \log(x/y) - \cdots + n \log(1 - \pi_{c-1}^*). \tag{3.6}$$

The term $\log(x/y)$ is a function of parameters $\{\alpha_j\}$ based on the proportional odds model. From the log likelihood function (3.6), the coefficients for the cumulative counts $\{X_j^*,\ j = 1, \ldots, c-1\}$ are the same. So, the likelihood function suggests that the cumulative counts contribute equally to the parameters. It does not hurt to treat them as independent samples. Therefore, this suggests that our proposed estimator naively treating all collapsings as independent samples has no loss of efficiency under the alternate probability case.

Table 3.1 summarizes some of the Pitman ARE results when the response has 3, 5, 7, or 9 categories. For the equal probability or alternate probability cases, we have ARE(MH,ML)=1; that is, our estimator has full asymptotic efficiency for testing the null hypothesis under both cases. The Pitman ARE for our estimator compared to the ML estimator is palindromic invariant (McCullagh, 1978), in the sense that the response probabilities $\pi_1 : \pi_2 : \cdots : \pi_c$ yield the same relative efficiency as $\pi_c : \pi_{c-1} : \cdots : \pi_1$. For example, the ARE(MH, ML)= 0.934 when the response probabilities have pattern 1 : 1 : 10 or 10 : 1 : 1. If there is only one response probability different from the others (*e. g.*, $p : p : p : q : p$), the asymptotic relative

efficiency for the proposed estimator tends to decrease as $|p - q|$ increases; that is, the proposed estimator tends to lose more efficiency as $|p - q|$ increases. If two response probabilities differ from the others (*e. g.*, $p : p : p : q_1 : q_2$), the asymptotic relative efficiency tends to increase as the position of $q_1 : q_2$ approaches the middle. In general, these is little loss of efficiency for the proposed estimator compared to the unconditional ML estimator in the large–stratum situation. In considering a wide range of cases with the permutation of '1', '4', and '7', the worst efficiency we noted was 0.813, when the probabilities have pattern $1 : 1 : 1 : 1 : 1 : 1 : 7 : 7 : 7$ (see, for some examples, Table 3.1).

Now, we generalize McCullagh's argument and study the Pitman ARE for the K strata case. First, we show that the efficient score test for the common proportional odds model (2.1) is equivalent to a test that generalize the Wilcoxon average rank test. The efficient score test is based on the log–likelihood derivative with respect to Δ evaluated at $\Delta = 0$ and $\alpha_{jk} = \hat{\alpha}_{jk}$, where $\hat{\alpha}_{jk}$ are the fitted parameters under the null hypothesis that $\Delta = 0$.

For the common proportional odds model (2.1), the cell probabilities can be written as follows:

$$\pi_{ijk} = \frac{\exp(\alpha_{jk} + \Delta I_{\{i=1\}})}{1 + \exp(\alpha_{jk} + \Delta I_{\{i=1\}})} - \frac{\exp(\alpha_{(j-1)k} + \Delta I_{\{i=1\}})}{1 + \exp(\alpha_{(j-1)k} + \Delta I_{\{i=1\}})} ,$$

where $i \in \{1, 2\}$, $j \in \{1, \ldots, c\}$, $k \in \{1, \ldots, K\}$, and $-\infty = \alpha_{0k} < \alpha_{1k} < \cdots < \alpha_{(c-1)k} < \alpha_{ck} = \infty$ for $k \in \{1, \ldots, K\}$. Denote the log–likelihood function by $\log L$,

which is proportional to $\sum_k \sum_i \sum_j X_{ijk} \log(\pi_{ijk})$. Then,

$$\frac{\partial \log L}{\partial \Delta}|_{\Delta=0} = \sum_k \{X_{11k} \frac{1}{\pi^0_{11k}} \pi^0_{11k}(1 - \pi^0_{11k})$$

$$+ \sum_{j=2}^{c-1} X_{1jk} \frac{1}{\pi^0_{1jk}} [\pi^{0*}_{1jk}(1 - \pi^{0*}_{1jk}) - \pi^{0*}_{1(j-1)k}(1 - \pi^{0*}_{1(j-1)k})]$$

$$+ (n_{1k} - X_{11k} - \cdots - X_{1(c-1)k}) \frac{1}{\pi^0_{1ck}} [-\pi^{0*}_{1(c-1)k}(1 - \pi^{0*}_{1(c-1)k})]\}, \qquad (3.7)$$

where $\{\pi^0_{1jk}\}$ are the fitted probabilities under the null hypothesis.

We now show that (3.7), which is the log–likelihood derivative with respect to Δ evaluated at $\Delta = 0$ and $\alpha_{jk} = \hat{\alpha}_{jk}$, equals

$$T = \sum_{k=1}^{K} \sum_{j=1}^{c} (n_{1k}\pi^0_{1jk} - X_{1jk})(\pi^{0*}_{1jk} + \pi^{0*}_{1(j-1)k}), \qquad (3.8)$$

where $\pi^{0*}_{10k} = 0$ for $k \in \{1, \ldots, K\}$. First, (3.7) simplifies to

$$\sum_k \{X_{11k}(1 - \pi^0_{11k}) + \sum_{j=2}^{c-1} X_{1jk} \frac{1}{\pi^0_{1jk}} [\pi^{0*}_{1jk} - \pi^{0*}_{1(j-1)k}$$

$$- (\pi^{0*}_{1jk} - \pi^{0*}_{1(j-1)k})(\pi^{0*}_{1jk} + \pi^{0*}_{1(j-1)k})] - (n_{1k} - X_{11k} - \cdots - X_{1(c-1)k})\pi^{0*}_{1(c-1)k}\}$$

$$= \sum_k \{X_{11k}(1 - \pi^0_{11k} + \pi^{0*}_{1(c-1)k}) + \sum_{j=2}^{c-1} X_{1jk} \frac{1}{\pi^0_{1jk}} [\pi^0_{1jk} - \pi^0_{1jk}(\pi^{0*}_{1jk} + \pi^{0*}_{1(j-1)k})]$$

$$+ (\sum_{j=2}^{c-1} X_{1jk})\pi^{0*}_{1(c-1)k} - n_{1k}(\pi^{0*}_{1(c-1)k})\}$$

$$= \sum_k [X_{11k}(1 - \pi^0_{11k} + \pi^{0*}_{1(c-1)k}) + \sum_{j=2}^{c-1} X_{1jk}(1 - \pi^{0*}_{1jk} - \pi^{0*}_{1(j-1)k} + \pi^{0*}_{1(c-1)k})$$

$$- n_{1k}(\pi^{0*}_{1(c-1)k})].$$

Also, since $\pi^0_{1ck} = 1 - \pi^0_{11k} - \cdots - \pi^0_{1(c-1)k}$ and $X_{1ck} = n_{1k} - X_{11k} - \cdots - X_{1(c-1)k}$, T simplifies to

$$\begin{aligned} T &= \sum_k [(n_{1k}\pi^0_{11k} - X_{11k})(\pi^{0*}_{11k}) + \sum_{j=2}^{c-1}(n_{1k}\pi^0_{1jk} - X_{1jk})(\pi^{0*}_{1jk} + \pi^{0*}_{1(j-1)k}) \\ &\quad - (n_{1k}\pi^0_{11k} - X_{11k} + \cdots + n_{1k}\pi^0_{1(c-1)k} - X_{1(c-1)k})(\pi^{0*}_{1ck} + \pi^{0*}_{1(c-1)k})] \,. \end{aligned}$$

This equals

$$\begin{aligned} &\sum_k [(n_{1k}\pi^0_{11k} - X_{11k})(\pi^{0*}_{11k} - \pi^{0*}_{1ck} - \pi^{0*}_{1(c-1)k}) \\ &+ \sum_{j=2}^{c-1}(n_{1k}\pi^0_{1jk} - X_{1jk})(\pi^{0*}_{1jk} + \pi^{0*}_{1(j-1)k} - \pi^{0*}_{1ck} - \pi^{0*}_{1(c-1)k})] \\ &= \sum_k \{X_{11k}(\pi^{0*}_{1ck} - \pi^{0*}_{11k} + \pi^{0*}_{1(c-1)k}) + \sum_{j=2}^{c-1} X_{1jk}(\pi^{0*}_{1ck} - \pi^{0*}_{1jk} - \pi^{0*}_{1(j-1)k} + \pi^{0*}_{1(c-1)k}) \\ &- n_{1k}[\pi^0_{11k}(\pi^{0*}_{1ck} - \pi^{0*}_{11k} + \pi^{0*}_{1(c-1)k}) + \sum_{j=2}^{c-1} \pi^0_{1jk}(\pi^{0*}_{1ck} - \pi^{0*}_{1jk} - \pi^{0*}_{1(j-1)k} + \pi^{0*}_{1(c-1)k})]\} \\ &= \sum_k [X_{11k}(1 - \pi^0_{11k} + \pi^{0*}_{1(c-1)k}) + \sum_{j=2}^{c-1} X_{1jk}(1 - \pi^{0*}_{1jk} - \pi^{0*}_{1(j-1)k} + \pi^{0*}_{1(c-1)k}) \\ &- n_{1k}(\pi^{0*}_{1(c-1)k} + \sum_{j=1}^{c-1} \pi^0_{1jk}(\pi^{0*}_{1(c-1)k} - \pi^{0*}_{1jk} - \pi^{0*}_{1(j-1)k}))] \,. \end{aligned}$$

The last term $\sum_{j=1}^{c-1} \pi^0_{1jk}(\pi^{0*}_{1(c-1)k} - \pi^{0*}_{1jk} - \pi^{0*}_{1(j-1)k})$ can be expanded as

$$\begin{aligned} &\pi^0_{11k}(\pi^0_{12k} + \cdots + \pi^0_{1(c-1)k}) + \pi^0_{12k}(\pi^0_{13k} + \cdots + \pi^0_{1(c-1)k} - \pi^0_{11k}) + \pi^0_{13k}(\pi^0_{14k} + \ldots \\ &+ \pi^0_{1(c-1)k} - \pi^0_{11k} - \pi^0_{12k}) + \cdots + \pi^0_{1(c-1)k}(-\pi^0_{11k} + \cdots + \pi^0_{1(c-2)k}) = 0 \,. \end{aligned}$$

So, T becomes

$$\begin{aligned} &\sum_k \; [X_{11k}(1 - \pi^0_{11k} + \pi^{0*}_{1(c-1)k}) + \sum_{j=2}^{c-1} X_{1jk}(1 - \pi^{0*}_{1jk} - \pi^{0*}_{1(j-1)k} + \pi^{0*}_{1(c-1)k}) - n_{1k}\pi^{0*}_{1(c-1)k}] \\ &\quad = \frac{\partial \log L}{\partial \Delta}|_{\Delta=0} \,. \end{aligned}$$

Therefore, the efficient score statistic and T are equivalent.

The term $n_{1k}\pi^0_{1jk} - X_{1jk}$ in T (3.8) is the residual under the null hypothesis for the first row and the jth column in the kth stratum. Similarly, we can consider the residual $n_{2k}\pi^0_{2jk} - X_{2jk}$ for the second row too. Also, the common proportional odds model (2.1) is equivalent to the model:

$$\log(\frac{\pi^*_{1jk}}{1-\pi^*_{1jk}}) = \alpha_{jk} + \frac{1}{2}\Delta\ , \quad \log(\frac{\pi^*_{2jk}}{1-\pi^*_{2jk}}) = \alpha_{jk} - \frac{1}{2}\Delta\ . \tag{3.9}$$

Following the same procedure as for model (2.1), the log–likelihood derivative with respect to Δ evaluated at the null for model (3.9), equals

$$T' = \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{c}(n_{1k}\pi^0_{1jk} - X_{1jk} - n_{2k}\pi^0_{1jk} + X_{2jk})(\pi^{0*}_{1jk} + \pi^{0*}_{1(j-1)k}), \tag{3.10}$$

where $\pi^{0*}_{10k} = 0$ for $k \in \{1,\ldots,K\}$.

Under the null hypothesis, $\pi^{0*}_{1jk} = m^*_{jk}/N_k$, where $m^*_{jk} = X^*_{1jk} + X^*_{2jk}$ is the marginal cumulative column total for the jth column in the kth stratum. Thus, a test based on the statistic T' (3.10) is equivalent to a test based on the statistic $T_W = \sum_{k=1}^{K}\sum_{j=1}^{c}(1/N_k)(X_{1jk} - X_{2jk})(m^*_{jk} + m^*_{(j-1)k} + 1)$, where $m^*_{0k} = 0$ for $k \in \{1,\ldots,K\}$. We call T_W a *weighted Wilcoxon average rank statistic.* If there is only one stratum, T_W is the Wilcoxon average rank statistic. Denote a *weighted Kendall's tau–a* by $\hat{\tau}_W = \sum_k(1/N_k)[\sum_{a=1}^{c-1} X_{1ak}\,(\sum_{b=a+1}^{c} X_{2bk}) - \sum_{b=1}^{c-1} X_{2bk}(\sum_{a=b+1}^{c} X_{1ak})]$. Similar to the one–stratum case, following Lehmann's argument (1975, p. 22), the weighted Wilcoxon average rank T_W equals $\hat{\tau}_W$ plus a constant, under fixed marginal totals n_{1k} and n_{2k} for $k \in \{1,\ldots,K\}$. Therefore, we can use an equivalent test based on the weighted Kendall's tau–a.

For the K strata case, the form of the Pitman ARE for our estimator compared to the unconditional ML estimator is similar to that for the one stratum case. So, the details of the calculations are not included. From Section 2.4, we know that $\sqrt{N_+}(\hat{\Delta} - \Delta)$ is $AN\,(0,\ (\sum_k t_k^2/\omega_k)/(\sum_k t_k)^2)$, where $\omega_k^{-1} = (1/a_{1k})[\sum_{a=1}^{c}\pi_{1ak}\eta^2_{1ak}] +$

$(1/a_{2k})[\sum_{b=1}^{c}\pi_{2bk}\eta_{2bk}^2]$, $t_k = (a_{1k}^{-1}+a_{2k}^{-1})^{-1}[\sum_{j=1}^{c-1}\pi_{2jk}^{*}(1-\pi_{1jk}^{*})]$, $\eta_{1ak} = (1/A_k)[\sum_{b=a+1}^{c}(b-a)\pi_{2bk}] - (1/B_k)[\sum_{b=1}^{a-1}(a-b)\pi_{2bk}]$, $\eta_{2bk} = (1/A_k)[\sum_{a=1}^{b-1}(b-a)\pi_{1ak}] - (1/B_k)[\sum_{a=b+1}^{c}(a-b)\pi_{1ak}]$, $A_k = \sum_{a<b}(b-a)\pi_{1ak}\pi_{2bk}$, $B_k = \sum_{a>b}(a-b)\pi_{1ak}\pi_{2bk}$, $a, b \in \{1, \ldots, c\}$, and $N_+a_{1k} = n_{1k}$, $N_+a_{2k} = n_{2k}$. Under the null hypothesis, $(\omega_k^0)^{-1} = (1/a_{1k} + 1/a_{2k})(1/A_k^0)^2[\sum_{a=1}^{c}\pi_{1ak}^0(\sum_{b=a+1}^{c}(b-a)\pi_{1bk}^0 - \sum_{b=1}^{a-1}(a-b)\pi_{1bk}^0)^2]$, $t_k^0 = (1/a_{1k} + 1/a_{2k})^{-1}[\sum_{j=1}^{c-1}\pi_{1jk}^{0*}(1-\pi_{1jk}^{0*})]$, and $A_k^0 = B_k^0 = \sum_{a<b}(b-a)\pi_{1ak}^0\pi_{1bk}^0 = \sum_{j=1}^{c}\pi_{1jk}^{0*}(1-\pi_{1jk}^{0*})$. The numerator of the efficacy of $\hat{\Delta}$ is $\mu'_{\hat{\Delta}}(\boldsymbol{\theta}_0) = 1$ and the denominator $N_+\sigma_{\hat{\Delta}}^2(\boldsymbol{\theta}_0)$ is

$$\sum_k (t_k^0)^2(\omega_k^0)^{-1}/(\sum_k t_k^0)^2 .$$

Therefore, the efficacy for the proposed estimator is

$$\mathrm{Eff}(\hat{\Delta}) = \frac{\sum_k (1/a_{1k} + 1/a_{2k})^{-1}[\sum_{j=1}^{c-1}\pi_{1jk}^{0*}(1-\pi_{1jk}^{0*})]}{\{\sum_k (1/a_{1k} + 1/a_{2k})^{-1}[\sum_{a=1}^{c}\pi_{1ak}^0(\sum_{b=a+1}^{c}(b-a)\pi_{1bk}^0 - \sum_{b=1}^{a-1}(a-b)\pi_{1bk}^0)^2]\}^{1/2}}.$$

Similarly, the efficacy of $\hat{\tau}_{\mathrm{W}}$ can be obtained by $\mu'_{\hat{\tau}_{\mathrm{W}}}(\boldsymbol{\theta}_0)/\sqrt{N_+\sigma_{\hat{\tau}_{\mathrm{W}}}^2(\boldsymbol{\theta}_0)}$. All calculations are similar to those in the one stratum case. We have

$$\mu'_{\hat{\tau}_{\mathrm{W}}}(\boldsymbol{\theta}_0) = \sum_k \frac{1}{a_{1k}+a_{2k}}(a_{1k}a_{2k})e_{1k} = \sum_k (\frac{1}{a_{1k}} + \frac{1}{a_{2k}})^{-1}e_{1k} ,$$

$$N_+\sigma_{\hat{\tau}_{\mathrm{W}}}^2(\boldsymbol{\theta}_0) = \sum_k \frac{1}{(a_{1k}+a_{2k})^2}(a_{1k}a_{2k})^2(\frac{1}{a_{1k}} + \frac{1}{a_{2k}})e_{2k} = \sum_k (\frac{1}{a_{1k}} + \frac{1}{a_{2k}})^{-1}e_{2k} ,$$

where

$$e_{1k} = \pi_{11k}^0(1-\pi_{11k}^0)(\pi_{11k}^0 + \pi_{12k}^0) + \sum_{a=2}^{c-1}\pi_{1ak}^{0*}(1-\pi_{1ak}^{0*})(\pi_{1(a+1)k}^0 + \pi_{1ak}^0)$$

$$e_{2k} = \pi_{11k}^0(\sum_{b=2}^{c}\pi_{1bk}^0)^2 + \sum_{a=2}^{c-1}\pi_{1ak}^0(\sum_{b=a+1}^{c}\pi_{1bk}^0 - \sum_{b=1}^{a-1}\pi_{1bk}^0)^2 + \pi_{1ck}^0(\sum_{b=1}^{c-1}\pi_{1bk}^0)^2 .$$

So, the Pitman ARE for the proposed estimator compared to the unconditional ML estimator for the general K strata case is

$$\text{ARE(MH, ML)} = \frac{[\sum_k (\frac{1}{a_{1k}} + \frac{1}{a_{2k}})^{-1} e^*_{1k}]^2}{\sum_k (\frac{1}{a_{1k}} + \frac{1}{a_{2k}})^{-1} e^*_{2k}} \Big/ \frac{[\sum_k (\frac{1}{a_{1k}} + \frac{1}{a_{2k}})^{-1} e_{1k}]^2}{\sum_k (\frac{1}{a_{1k}} + \frac{1}{a_{2k}})^{-1} e_{2k}}, \qquad (3.11)$$

where

$$e^*_{1k} = \sum_{j=1}^{c-1} \pi^{0*}_{1jk}(1 - \pi^{0*}_{1jk}),$$

$$e^*_{2k} = \sum_{a=1}^{c} \pi^0_{1ak}\Big(\sum_{b=a+1}^{c} (b-a)\pi^0_{1bk} - \sum_{b=1}^{a-1} (a-b)\pi^0_{1bk}\Big)^2.$$

Table 3.2 shows the Pitman ARE for the proposed estimator compared to the unconditional ML estimator when the response has 3 categories and the number of strata is larger than 1. If all strata have the same pattern of response probabilities, the term e^*_{1k} in the expression (3.11) is the same for all $k = 1, \ldots, K$, as well as $\{e^*_{2k}\}$, $\{e_{1k}\}$, and $\{e_{2k}\}$. Therefore, under the same pattern of response probabilities for all strata, the Pitman ARE from the expression (3.11) is the same as the one stratum situation. For example, when the response probability has pattern $1:1:10$ for all strata, ARE(MH, ML)=0.934, which is the same as the corresponding entry in Table 3.1.

For the one stratum case, alternating probabilities give full efficiency. However, for the two strata situation, if both strata have alternating probabilities but patterns differ from stratum to stratum, the proposed estimator has a loss of efficiency. For instance, when one stratum has pattern $1:1:1$ and another stratum has pattern $2:9:2$ and the sample sizes are equal for all rows, the Pitman ARE(MH, ML) is 0.987. If we let the sample sizes among the strata vary, the asymptotic relative efficiency does not change a lot. For the same example above, one sample size structure yields the Pitman ARE of 0.987 and another sample size structure yields the Pitman ARE of 0.986. For the three strata case, the results are similar to those for the two strata case.

For the general K strata case, the loss of efficiency for the proposed estimator depends on the probability pattern for each stratum. In general, the proposed estimator does not have much loss of efficiency compared to the unconditional ML estimator.

Table 3.1. The Pitman ARE for the proposed estimator compared to the unconditional ML estimator

c	$\pi_1 : \pi_2 : \cdots : \pi_c$	ARE(MH, ML)	$\pi_1 : \pi_2 : \cdots : \pi_c$	ARE(MH, ML)
3	1 : 1 : 1	1.000	2 : 9 : 2	1.000
	1 : 1 : 2	0.990	2 : 4 : 5	0.986
	1 : 1 : 10	0.934	10 : 1 : 1	0.934
5	1 : 1 : 1 : 1 : 1	1.000	1 : 5 : 1 : 5 : 1	1.000
	3 : 1 : 5 : 3 : 7	0.972	1 : 3 : 1 : 3 : 7	0.955
	1 : 1 : 1 : 5 : 7	0.860	1 : 1 : 5 : 7 : 1	0.919
	1 : 5 : 7 : 1 : 1	0.932	5 : 7 : 1 : 1 : 1	0.866
	1 : 1 : 1 : 1 : 5	0.953	1 : 1 : 1 : 5 : 1	0.942
	1 : 1 : 5 : 1 : 1	0.968	1 : 1 : 3 : 1 : 1	0.985
7	1 : 1 : 1 : 1 : 1 : 1 : 1	1.000	1 : 4 : 1 : 4 : 1 : 4 : 1	1.000
	1 : 4 : 1 : 1 : 7 : 1 : 7	0.976	1 : 7 : 7 : 7 : 4 : 1 : 4	0.956
	1 : 1 : 1 : 1 : 1 : 4 : 7	0.864	1 : 1 : 1 : 1 : 4 : 7 : 1	0.869
	1 : 1 : 1 : 4 : 7 : 1 : 1	0.906	1 : 1 : 4 : 7 : 1 : 1 : 1	0.916
	1 : 4 : 7 : 1 : 1 : 1 : 1	0.880	4 : 7 : 1 : 1 : 1 : 1 : 1	0.860
	1 : 1 : 1 : 1 : 1 : 1 : 4	0.974	1 : 1 : 1 : 1 : 1 : 4 : 1	0.958
	1 : 1 : 1 : 1 : 4 : 1 : 1	0.967	1 : 1 : 1 : 4 : 1 : 1 : 1	0.975

c	$\pi_1 : \pi_2 : \cdots : \pi_c$	ARE(MH, ML)
9	1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 4	0.981
	1 : 1 : 1 : 1 : 1 : 1 : 1 : 4 : 1	0.967
	1 : 1 : 1 : 1 : 1 : 1 : 4 : 1 : 1	0.968
	1 : 1 : 1 : 1 : 1 : 4 : 1 : 1 : 1	0.975
	1 : 1 : 1 : 1 : 4 : 1 : 1 : 1 : 1	0.978
	1 : 1 : 1 : 1 : 1 : 1 : 1 : 4 : 7	0.879
	1 : 1 : 1 : 1 : 1 : 1 : 4 : 7 : 1	0.869
	1 : 1 : 1 : 1 : 1 : 4 : 7 : 1 : 1	0.885
	1 : 1 : 1 : 1 : 4 : 7 : 1 : 1 : 1	0.908

Table 3.2. The Pitman ARE for the proposed estimator compared to the ML estimator under the K strata and 3 response categories case.

K	a_{ik}	$\pi_{111} : \pi_{121} : \pi_{131}$	$\pi_{112} : \pi_{122} : \pi_{132}$	$\pi_{113} : \pi_{123} : \pi_{133}$	ARE(MH, ML)
2	I	1:1:1	1:1:1	N.A.	1.00
2	I	2:9:2	2:9:2	N.A.	1.00
2	I	1:1:10	1:1:10	N.A.	0.934
2	II	1:1:1	2:9:2	N.A.	0.987
	III	1:1:1	2:9:2	N.A.	0.986
2	II	1:1:1	2:8:2	N.A.	0.988
	III	1:1:1	2:8:2	N.A.	0.988
2	II	1:1:1	1:1:10	N.A.	0.977
	III	1:1:1	1:1:10	N.A.	0.962
3	II	1:1:1	2:9:2	1:1:1	0.990
	IV	1:1:1	2:9:2	1:1:1	0.992
3	II	1:1:1	2:9:2	2:9:2	0.986

Note: In the column a_{ik}, we use I–IV to represent the different sample sizes structures, where I: a_{ik} is arbitrary; II: a_{ik} is equal for $i \in \{1,2\}$, $k \in \{1,\ldots,K\}$; III: $a_{11} = 0.1$, $a_{21} = 0.2$, $a_{12} = 0.3$, $a_{22} = 0.4$; and IV: $a_{11} = 0.1$, $a_{21} = 0.2$, $a_{12} = 0.1$, $a_{22} = 0.1$, $a_{13} = 0.1$, $a_{23} = 0.4$. The N.A. indicates 'not applicable'.

3.3 Simulation Study

The previous section illustrates that in the large–stratum case, there is little loss of efficiency for the proposed estimator compared to the unconditional ML estimator, near conditional independence. In this section, we discuss a simulation study that investigates the robustness of our estimator and the ML estimator, particularly as a function of the sparseness of the data. The results show that the proposed estimator is much better than the unconditional ML estimator under the sparse–data situation.

This study uses the simpler proportional odds model (2.25), since software for ML estimation is widely available for it. For simplicity, we assign equal probabilities for the c–category response in the second row for each $2 \times c$ contingency table, assuming a lack of stratum effects (*i. e.*, $\beta_k = 0$ for all k). Under model (2.25), the probabilities in the first row for each partial table depend only on the common cumulative odds ratio ψ. We take $c = 3$, 5, or 7, and K =5, 10, or 25. For each combination of c and K, we choose ψ =1, 2, or 4, and for each k, $n_{1k} = n_{2k} = 20$, $n_{1k} = n_{2k} = 1$, $n_{1k} = 2$, $n_{2k} = 3$, $n_{1k} = 1$, $n_{2k} = 10$ or $n_{1k} = 10$, $n_{2k} = 1$. The reason that we also switch the sample sizes for the rows in the highly unbalanced cases is because equal probabilities are assigned only in the second row. Therefore, results may differ according to whether the sole observation falls in the first row or the second row.

The combination of the sample sizes provides a variety of scenarios, ranging from ones for which large-strata asymptotics should work well (*e. g.*, $K = 5$ and $n_{ik} = 20$) to ones for which sparse-data limiting asymptotics would seem more appropriate (*e. g.*, $K = 25$ and $n_{ik} \leq 3$), as well as ones with unbalanced sample sizes. For each case, we generate 10,000 $2 \times c \times K$ sample tables based on sampling independent multinomial distributions, in order to estimate various characteristics of the ML and MH estimators of the proportional odds parameter.

Our random generation of multinomial distributions utilizes the uniform distribution. First, we generate a random number z from a uniform distribution over the interval (0, 1). If z is greater than π^*_{j-1} but less than π^*_j, the first observation falls in the jth column. We repeat this step n times. The number of times for which an observation falls in the jth column is the generated sample count for the jth column. These sample counts follow a multinomial distribution with total sample size n.

For the 10,000 simulated tables for each case, we calculated the sample mean, the mean squared error, and the mean absolute deviation for $\hat{\Delta}$, for the unconditional ML estimator and the proposed Mantel–Haenszel (MH) type estimator. Table 3.3 summarizes the sample means for the unconditional ML estimator and for the MH estimator. Standard errors for the sample mean are measures of the precision the 10,000 simulations provide. These standard errors for both the ML estimator and the MH estimator and the correlation between them are used to determine when a significant difference occurs between them. The standard errors are not shown in the tables. Briefly, most of the standard errors for the estimators are less than 0.01. However, to compare the ML and MH estimators in Table 3.3, a * symbol indicates that that estimate is significantly better than the other estimate based on the measures of the precision and their correlation coefficients for 10,000 simulations, when the type I error probability is 0.05. Here, we mean that estimate 'A' is significantly better than estimate 'B' if 'A' is significantly closer to the true parameter.

Table 3.3 shows that the MH estimator tends to be closer to the true parameter than the ML estimator, since there are many more *'s for the MH estimates in the table. Especially, for the sparse–data case ($n_{ik} \leq 3$), the MH estimator is much better than the ML estimator. In general, when K increases, the MH estimate tends to be closer to the true parameter. This is because when the number of observations increases, we have more information to estimate parameters. However, for the ML estimate, this property fails when data are extremely sparse. When $n_{1k} = n_{2k} = 1$,

the ML estimate does not seem to be bad for the $K = 5$ case but it seems to be worse for the $K = 10$ and $k = 25$ cases. When $n_{1k} = 2$ and $n_{2k} = 3$, the ML estimate seems to be worse compared to the MH estimate, but within the ML method, it tends to be better when K increases. Also, in general, the ML estimate tends to be closer to the true parameter when the number of response categories increases, except for the case with extremely few observations (*e. g.*, $K = 5$ and $n_{1k} = n_{2k} = 1$). Intuitively, when the number of response categories increases, the response tends toward a continuous variable. Therefore, we have more information to specify the association for the two groups.

For the special case of $n_{1k} = n_{2k} = 1$ and $K = 10$ or 25, the ML estimate of Δ is about twice the true parameter. Under the same situation, the MH estimator is much closer to the true parameter. When $K = 5$ and $n_{1k} = n_{2k} = 1$, both the ML and the MH estimates are not close to the true parameter. This case is not surprising, because we cannot expect a good estimate based on only a few observations. For the unbalanced sample sizes ($n_{1k} = 1$ and $n_{2k} = 10$, or $n_{1k} = 10$ and $n_{2k} = 1$), there is not much difference between the MH and the ML estimates when K is small; but when K is large, the MH estimate is significantly better than the ML estimate. Also, the estimates seem to be better for the case $n_{1k} = 10$ and $n_{2k} = 1$ than for the case $n_{1k} = 1$ and $n_{2k} = 10$. Intuitively, when the first row has more observations, we can specify the association clearly because the probabilities assigned in the first row depend on ψ and the probabilities assigned in the second row are equal.

Table 3.4 shows the mean squared errors for those two estimators. Tables 3.5 shows the mean absolute deviation, $E|\hat{\Delta} - \Delta|$, for two estimators. The mean absolute deviation is an alternative way to the mean squared error to measure the closeness of an estimator to a parameter. Standard errors for the mean squared error or the mean absolute deviation are measures of the precision the 10,000 simulations provide. These standard errors for both the ML method and the MH method and the correlation

between them are used to determine when a significant difference occurs between them. As in the sample mean case, in Tables 3.4 or 3.5 the * symbol indicates that the mean squared error or the mean absolute deviation for that estimate is significantly smaller than that for the other estimate, when the type I error probability is 0.05. Most of the standard errors for the mean squared errors or the mean absolute deviations are less than 0.01, except for the cases with high deviations (such as for the ML method in the sparse–data case). Both tables yields similar results as the sample mean for comparing the two estimators. When the data are sparse, the MH estimators have significantly smaller mean squared errors than the ML estimators. For the unbalanced cases, the ML estimate is better than the MH estimate when $K = 5$, but the MH estimate is better than the ML estimate when $K = 10$ or 25.

Table 3.6 shows the proportion of times that the ML estimator was closer than the MH estimator to the true Δ. Since the proportion is based on binomial sampling with sample size 10,000, its largest possible standard error is $(0.5 \times 0.5/10000)^{0.5} = 0.005$. For the large–strata situation (*e. g.*, $n_{1k} = n_{2k} = 20$), there is no large difference between those two estimators. On the other hand, for the sparse–data situation (*e. g.*, $n_{1k} = n_{2k} = 1$, or $n_{1k} = 2$ and $n_{2k} = 3$), the MH estimator obviously dominates the ML estimator. Under conditional independence ($\psi = 1$) and for sample sizes $n_{1k} = n_{2k} = 1$, the MH estimator is preferable for almost all cases, especially when the number of response categories equals 3. For the unbalanced sample cases, there is no large difference between the two estimators. However, in these cases, the proportion of times that the ML estimator is closer than the MH estimator to Δ decreases as the number of strata increases.

Furthermore, we analyze the asymptotic approximations in the upper tail for these estimators. Table 3.7 reports sample proportions estimating the probability that the Wald statistics based on these estimates exceed the $100(1 - \alpha)$ percentage point of the relevant chi–squared distribution. Similar to the proportion of times in Table

3.6, the largest standard error of the sample proportions is 0.005. When the sample proportion is between 0.05 and 0.1, its standard error is between 0.002 and 0.003. For small sample sizes within each stratum, the ML statistics are highly liberal, especially when the number of strata is large. The MH statistics behave very well except for the case when the number of strata equals 5 and the sample sizes are $n_{1k} = n_{2k} = 1$. Next, we explain why the size for this statistic equals zero under $\alpha = 0.05$ and $c = 3$.

If we consider the ordinary Mantel–Haenszel estimator for $2 \times 2 \times K$ contingency tables, the Mantel–Haenszel test with $n_{1k} = n_{2k} = 1$ can be viewed as identical to McNemar's test (Agresti, 1990, p. 394). For instance, for five strata, it corresponds to having five observations in the two off–diagonal cells of the 2×2 contingency table for the McNemar representation of the data. For five observations, the exact binomial analysis could at best achieve a p–value of $1/16 = 0.0625$ for a two–sided alternative, so we should also hope to never reject with an asymptotic test when the type I error probability is 0.05. For the three response categories case, Section 2.6 presented the simpler MH estimators for the special case with one observation in each row, using a $c \times c$ squared table. When $c = 3$, $K = 5$ and $n_{1k} = n_{2k} = 1$, we consider the 3×3 squared table having five observations in the six off–diagonal cells. The extreme situation is when $x_{13} = 4$ and $x_{21} = 1$. Notice that the situation of $x_{13} = 5$ is not considered, since the cumulative odds ratio estimate for that case will be infinity. For the extreme situation, the log MH estimate is 2.079 and its standard error is 1.118; the best p-value achieved for testing the null based on the Wald statistic, is then 0.063. Therefore, we never reject the null when the type I error probability is 0.05. When the number of the response categories increases, the test becomes more powerful. The case $c = 5$ has fewer problems than $c = 3$, since the size is 0.014 when $c = 5$ but the size is 0.0000 when $c = 3$. Also, for unbalanced cases, the ML estimator is more liberal when $K = 10$ or 25.

In general, the two estimators are almost the same under the large–stratum situation, but the proposed Mantel–Haenszel type estimator tends to be much better than the ML estimator under the sparse–data situation. For the unbalanced cases, when K is large, the proposed estimator is better than the ML estimator. In addition, both estimators are poor if there are a few observations.

Table 3.3. The sample mean for $\hat{\Delta}$ over 10,000 simulations for both the unconditional ML and the proposed Mantel–Haenszel (MH) estimates

			$c = 3$		$c = 5$		$c = 7$	
			$\psi = 2$	$\psi = 4$	$\psi = 2$	$\psi = 4$	$\psi = 2$	$\psi = 4$
	K	n_{1k}, n_{2k}	$\Delta = 0.693$	1.386	0.693	1.386	0.693	1.386
ML	5	20,20	0.712	1.429	0.710	1.426	0.712	1.425
		1,1	0.720*	1.294*	1.149	1.857	0.997	1.790
		2,3	0.900	1.727	0.878	1.711	0.870	1.705
		1,10	0.742*	1.357	0.770*	1.492*	0.764*	1.514*
		10,1	0.751	1.523	0.753	1.520	0.757	1.524
MH	5	20,20	0.700*	1.404*	0.700*	1.405*	0.702*	1.406*
		1,1	0.392	0.693	0.511*	0.934*	0.566*	1.049*
		2,3	0.784*	1.492*	0.783*	1.524*	0.780*	1.536*
		1,10	0.768	1.411	0.824	1.612	0.827	1.652
		10,1	0.735*	1.452*	0.738*	1.479*	0.744*	1.492*
ML	10	20,20	0.712	1.424	0.710	1.419	0.710	1.418
		1,1	1.241	2.310	1.254	2.428	1.234	2.400
		2,3	0.867	1.740	0.837	1.681	0.828	1.667
		1,10	0.786	1.549	0.760	1.532*	0.753	1.516*
		10,1	0.750	1.529	0.746	1.504	0.745	1.498
MH	10	20,20	0.699*	1.398*	0.699*	1.396*	0.699*	1.387*
		1,1	0.666*	1.200*	0.742*	1.372*	0.759*	1.432*
		2,3	0.740*	1.482*	0.731*	1.470*	0.724*	1.470*
		1,10	0.768*	1.518*	0.757	1.540	0.756	1.533
		10,1	0.711*	1.429*	0.715*	1.435*	0.716*	1.436*
ML	25	20,20	0.708	1.418	0.706	1.413	0.706	1.412
		1,1	1.362	2.847	1.201	2.515	1.146	2.373
		2,3	0.836	1.694	0.812	1.640	0.805	1.622
		1,10	0.756	1.533	0.742	1.495	0.739	1.485
		10,1	0.748	1.520	0.740	1.490	0.738	1.482
MH	25	20,20	0.694*	1.390*	0.694*	1.389*	0.695*	1.390*
		1,1	0.752*	1.486*	0.745*	1.491*	0.742*	1.492*
		2,3	0.706*	1.421*	0.702*	1.412*	0.702*	1.411*
		1,10	0.715*	1.450*	0.710*	1.435*	0.711*	1.432*
		10,1	0.700*	1.403*	0.700*	1.404*	0.701*	1.404*

Table 3.4. The mean squared error for $\hat{\Delta}$ over 10,000 simulations for both the unconditional ML and the proposed Mantel–Haenszel (MH) estimates

			$c = 3$		$c = 5$		$c = 7$	
			$\psi = 2$	$\psi = 4$	$\psi = 2$	$\psi = 4$	$\psi = 2$	$\psi = 4$
	K	n_{1k}, n_{2k}	$\Delta = 0.693$	1.386	0.693	1.386	0.693	1.386
ML	5	20,20	0.075	0.088	0.069	0.079	0.068	0.077
		1,1	2.333	1.962	18.301	13.570	5.948	4.966
		2,3	1.062	1.206	0.948	1.122	0.928	1.077
		1,10	0.931*	0.750*	0.919*	0.897*	0.888*	0.926*
		10,1	1.001*	1.087	0.923*	0.997*	0.890*	0.974*
MH	5	20,20	0.072*	0.086*	0.067*	0.078	0.066*	0.076
		1,1	0.776*	1.048*	1.018*	1.028*	1.138*	1.089*
		2,3	0.782*	0.834*	0.744*	0.870*	0.731*	0.884*
		1,10	1.053	0.969	1.122	1.320	1.104	1.411
		10,1	1.074	1.093	1.011	1.077	0.992	1.076
ML	10	20,20	0.037	0.044	0.034	0.039	0.034	0.038
		1,1	2.965	3.295	2.969	4.041	2.769	3.672
		2,3	0.512	0.724	0.434	0.583	0.418	0.547
		1,10	0.535	0.615*	0.455*	0.537*	0.435*	0.497*
		10,1	0.492	0.555	0.427	0.461	0.413	0.447
MH	10	20,20	0.036*	0.042*	0.033*	0.038*	0.032*	0.037*
		1,1	0.741*	0.632*	0.838*	0.744*	0.870*	0.824*
		2,3	0.358*	0.486*	0.318*	0.432*	0.310*	0.409*
		1,10	0.534	0.702	0.477	0.691	0.459	0.654
		10,1	0.461*	0.503*	0.410*	0.448*	0.397*	0.435*
ML	25	20,20	0.015	0.018	0.014	0.016	0.014	0.015
		1,1	1.719	4.210	1.149	2.655	0.978	2.101
		2,3	0.200	0.320	0.170	0.247	0.163	0.228
		1,10	0.193	0.264	0.167	0.200*	0.162	0.184*
		10,1	0.188	0.219	0.168	0.183	0.162	0.177
MH	25	20,20	0.014*	0.017*	0.013*	0.015*	0.013*	0.015*
		1,1	0.368*	0.468*	0.332*	0.451*	0.317*	0.448*
		2,3	0.129*	0.168*	0.117*	0.146*	0.114*	0.141*
		1,10	0.176*	0.256*	0.159*	0.214	0.155*	0.205
		10,1	0.165*	0.179*	0.150*	0.161*	0.146*	0.157*

Table 3.5. Sample mean absolute deviation, $\hat{E}|\hat{\Delta} - \Delta|$ over 10,000 simulations for both the unconditional ML and the proposed Mantel–Haenszel (MH) estimates

			$c = 3$		$c = 5$		$c = 7$	
			$\psi = 2$	$\psi = 4$	$\psi = 2$	$\psi = 4$	$\psi = 2$	$\psi = 4$
	K	n_{1k}, n_{2k}	$\Delta = 0.693$	1.386	0.693	1.386	0.693	1.386
ML	5	20,20	0.217	0.234	0.208	0.224	0.207	0.222
		1,1	1.262	1.126	1.842	1.623	1.523	1.430
		2,3	0.801	0.858	0.756	0.815	0.748	0.823
		1,10	0.771*	0.712*	0.753*	0.759*	0.739*	0.759*
		10,1	0.790*	0.822	0.751*	0.778*	0.737*	0.765*
MH	5	20,20	0.214*	0.232*	0.205*	0.223	0.204*	0.221
		1,1	0.698*	0.784*	0.809*	0.799*	0.850*	0.833*
		2,3	0.687*	0.725*	0.662*	0.720*	0.656*	0.720*
		1,10	0.798	0.792	0.803	0.895	0.797	0.911
		10,1	0.796	0.810*	0.767	0.794	0.757	0.787
ML	10	20,20	0.154	0.166	0.148	0.157	0.146	0.155
		1,1	1.351	1.458	1.285	1.503	1.245	1.453
		2,3	0.557	0.658	0.517	0.593	0.508	0.574
		1,10	0.570	0.616*	0.527	0.570*	0.518*	0.548*
		10,1	0.553	0.584	0.515	0.535	0.508	0.526
MH	10	20,20	0.151*	0.164*	0.145*	0.156*	0.144*	0.154*
		1,1	0.684*	0.642*	0.715*	0.697*	0.721*	0.724*
		2,3	0.465*	0.535*	0.441*	0.506*	0.436*	0.491*
		1,10	0.559*	0.639	0.529	0.626	0.523	0.611
		10,1	0.530*	0.550*	0.501*	0.521*	0.495*	0.515*
ML	25	20,20	0.098	0.107	0.094	0.101	0.093	0.099
		1,1	0.994	1.590	0.828	1.261	0.767	1.127
		2,3	0.350	0.443	0.325	0.392	0.318	0.377
		1,10	0.347	0.400	0.324	0.353*	0.320	0.340*
		10,1	0.345	0.371	0.326	0.340	0.321	0.334
MH	25	20,20	0.095*	0.103*	0.092*	0.098*	0.091*	0.097*
		1,1	0.471*	0.533*	0.450*	0.514*	0.442*	0.509*
		2,3	0.282*	0.321*	0.269*	0.300*	0.266*	0.295*
		1,10	0.331*	0.389*	0.314*	0.362	0.312*	0.354
		10,1	0.322*	0.334*	0.309*	0.319*	0.304*	0.315*

Table 3.6. Proportion of times that the unconditional ML estimate is closer than the Mantel–Haenszel–type estimator to the true parameter; that is, the proportion of times that $|\hat{\Delta}(\mathrm{ML}) - \Delta| < |\hat{\Delta}(\mathrm{MH}) - \Delta|$, over 10,000 simulations

			$c = 3$			$c = 5$			$c = 7$	
K	n_{1k}, n_{2k}	$\psi = 1$	2	4	1	2	4	1	2	4
5	20,20	.402	.457	.499	.423	.468	.513	.426	.478	.508
	1,1	.004	.145	.355	.143	.209	.313	.174	.229	.317
	2,3	.273	.348	.427	.310	.370	.448	.325	.380	.451
	1,10	.500	.538	.605	.534	.547	.596	.536	.550	.599
	10,1	.507	.516	.525	.531	.534	.548	.539	.541	.543
10	20,20	.400	.462	.493	.417	.465	.499	.422	.472	.505
	1,1	.019	.199	.309	.113	.203	.290	.141	.220	.285
	2,3	.223	.340	.414	.287	.358	.434	.298	.367	.424
	1,10	.429	.487	.562	.462	.499	.569	.469	.505	.571
	10,1	.428	.467	.484	.460	.489	.512	.472	.494	.515
25	20,20	.389	.460	.468	.413	.461	.478	.413	.456	.486
	1,1	.036	.256	.225	.106	.262	.250	.132	.271	.259
	2,3	.196	.383	.364	.267	.377	.376	.285	.376	.379
	1,10	.372	.470	.520	.412	.491	.547	.409	.489	.554
	10,1	.366	.442	.464	.406	.454	.483	.418	.462	.488

Table 3.7. Sample proportions estimating the probability that the Wald statistics based on these estimates exceeded the $100(1-\alpha)$ percentage point of the relevant chi-squared distribution

			$c=3$		$c=5$		$c=7$	
	K	n_{1k}, n_{2k}	$\alpha=0.05$	0.10	0.05	0.10	0.05	0.10
ML	5	20,20	0.056	0.106	0.057	0.110	0.057	0.107
		1,1	0.013	0.083	0.048	0.143	0.072	0.179
		2,3	0.075	0.143	0.088	0.160	0.091	0.162
		1,10	0.034	0.098	0.061	0.120	0.068	0.125
		10,1	0.036	0.095	0.061	0.120	0.068	0.126
MH	5	20,20	0.053	0.101	0.053	0.104	0.053	0.103
		1,1	0.000	0.003	0.014	0.067	0.041	0.101
		2,3	0.046	0.105	0.056	0.112	0.056	0.113
		1,10	0.043	0.088	0.050	0.108	0.053	0.115
		10,1	0.046	0.090	0.052	0.108	0.056	0.109
ML	10	20,20	0.055	0.104	0.056	0.109	0.056	0.107
		1,1	0.142	0.235	0.187	0.274	0.201	0.290
		2,3	0.080	0.146	0.087	0.155	0.090	0.154
		1,10	0.057	0.114	0.065	0.118	0.068	0.124
		10,1	0.057	0.113	0.061	0.116	0.064	0.123
MH	10	20,20	0.052	0.099	0.052	0.104	0.052	0.102
		1,1	0.017	0.073	0.060	0.115	0.068	0.122
		2,3	0.048	0.099	0.052	0.103	0.053	0.101
		1,10	0.045	0.094	0.049	0.100	0.051	0.104
		10,1	0.041	0.097	0.045	0.100	0.048	0.099
ML	25	20,20	0.053	0.107	0.053	0.105	0.052	0.107
		1,1	0.181	0.266	0.191	0.274	0.189	0.275
		2,3	0.084	0.148	0.091	0.154	0.092	0.156
		1,10	0.062	0.117	0.062	0.115	0.065	0.117
		10,1	0.064	0.126	0.064	0.124	0.064	0.126
MH	25	20,20	0.049	0.101	0.049	0.100	0.049	0.101
		1,1	0.045	0.099	0.050	0.104	0.052	0.104
		2,3	0.048	0.099	0.052	0.103	0.051	0.101
		1,10	0.045	0.097	0.047	0.095	0.049	0.096
		10,1	0.049	0.103	0.049	0.101	0.050	0.104

3.4 Tests of the Common Cumulative Odds Ratio Assumption

We first discuss tests of the common odds ratio for $2 \times 2 \times K$ contingency tables. A possible hypothesis test of homogeneous odds ratios across the strata can be conducted by testing the fit of the additive logit model with the ML method, provided the within–stratum sample sizes are adequate. The additive logit model is a special case of the common proportional odds model (2.1) when the response has two categories. The test is equivalent to a test of no interaction between the explanatory variable Z and the control variable W. The test statistics have asymptotically chi–squared distributions with $K-1$ degrees of freedom (df), when the additive logit model holds.

Another possible test was provided by Breslow and Day (1980). They developed a test for this homogeneous odds ratio hypothesis that does not require fitting logit models. The Breslow–Day test statistic is focused directly at the potential lack of homogeneity of odds ratios. It sums up the squared deviations of observed and fitted values, each standardized by its variance,

$$Q_{\mathrm{BD}} = \sum_{k=1}^{K} \frac{[X_{11k} - \mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})]^2}{\mathrm{Var}(X_{11k}|\hat{\psi}_{\mathrm{MH}})}. \tag{3.12}$$

In the Breslow and Day test statistic (3.12), $\mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})$ is the expected value of X_{11k} under the homogeneous odds ratios assumption. The variance $\mathrm{Var}(X_{11k}|\hat{\psi}_{\mathrm{MH}})$ they used in (3.12) is an asymptotic variance from the asymptotically normal distribution. Both mean and variance are calculated by treating $\hat{\psi}_{\mathrm{MH}}$ (2.2) as the true odds ratio. When a particular stratum has an odds ratio that is much larger or smaller than $\hat{\psi}_{\mathrm{MH}}$, we expect the distance between the observed value X_{11k} and the fitted value $\mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})$ to be large. The statistic Q_{BD} seems a reasonable test for the adequacy of the assumption of a common odds ratio when the sample size within each stratum is large. We next discuss the details of the mean and variance.

For $2 \times 2 \times K$ contingency tables, given the row and column marginal totals, $X_{12k} = n_{1k} - X_{11k}$, $X_{21k} = m_{1k} - X_{11k}$, and $X_{22k} = n_{2k} - m_{1k} + X_{11k}$. The expected values $\mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})$ can be solved from the quadratic equation:

$$\hat{\psi}_{\mathrm{MH}} = \frac{\mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})(n_{2k} - m_{1k} + \mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}}))}{(n_{1k} - \mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}}))(m_{1k} - \mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}}))}. \tag{3.13}$$

One of the roots of the quadratic equation yields a possible value for $\mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})$ in the sense that the expected values of X_{11k}, X_{12k}, X_{21k}, X_{22k} are all positive. The asymptotic variance of X_{11k} given $\hat{\psi}_{\mathrm{MH}}$ is given by the reciprocals of the fitted values:

$$\begin{aligned}\mathrm{Var}(X_{11k}|\hat{\psi}_{\mathrm{MH}}) &= \left[\frac{1}{\mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})} + \frac{1}{n_{1k} - \mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})}\right. \\ &\quad \left. + \frac{1}{m_{1k} - \mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})} + \frac{1}{n_{2k} - m_{1k} + \mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})}\right]^{-1}.\end{aligned}$$

The statistic Q_{BD} (3.12) also can be written as

$$\begin{aligned}Q_{\mathrm{BD}} &= \sum_{k=1}^{K}[X_{11k} - \mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})]^2\left[\frac{1}{\mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})} + \frac{1}{n_{1k} - \mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})}\right. \\ &\quad \left. + \frac{1}{m_{1k} - \mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})} + \frac{1}{n_{2k} - m_{1k} + \mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})}\right] \\ &= \sum_{k=1}^{K}\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{[X_{ijk} - \mathrm{E}(X_{ijk}|\hat{\psi}_{\mathrm{MH}})]^2}{\mathrm{E}(X_{ijk}|\hat{\psi}_{\mathrm{MH}})},\end{aligned}$$

since the term $[X_{ijk} - \mathrm{E}(X_{ijk}|\hat{\psi}_{\mathrm{MH}})]^2$ is constant for $i = 1, 2$ and $j = 1, 2$, given the fixed margins. Therefore, the Breslow–Day test statistic is a Pearson chi-square statistic with expected values from (3.13). If the unconditional ML estimate of the logit model is inserted in the Breslow–Day statistic (3.12) instead of $\hat{\psi}_{\mathrm{MH}}$, the statistic is then a Pearson chi-square goodness-of-fit statistic for the logit model (Breslow and Day, 1980, p. 142). The statistic has asymptotically a chi-squared distribution with df= $K - 1$ when the model holds.

Breslow and Day (1980) claimed that their test statistic Q_{BD} (3.12) follows approximately a chi–squared distribution with df= $K-1$ if the homogeneity assumption is valid and the sample size within each stratum is large. However, Tarone (1985) found that Q_{BD} (3.12) is not asymptotically a chi–squared statistic, and he developed a correction method so that the new statistic follows an asymptotically chi–squared distribution. He claimed that the Breslow and Day test is equivalent to the heterogeneity score test based on the conditional likelihood function. The idea in Tarone's arguments is that if an inefficient estimator (such as $\hat{\psi}_{\text{MH}}$) is used in computing the heterogeneity score test statistic, then the score test will be stochastically larger under the homogeneous odds ratio hypothesis than a chi–squared random variable with $K-1$ degrees of freedom. So, the Breslow and Day test is a liberal test since we reject more often than we should.

Next, we introduce some possible tests for the common cumulative odds ratio assumption for the ordinal response case. A proportional odds model assumes that the proportional odds assumption holds; that is, the odds ratio for each collapsing is identical. In our case with the common proportional odds model (2.1), the odds ratio is constant for all collapsings and all strata. Since the proposed estimator is based on the common proportional odds model, we may consider a test of homogeneous odds ratios across the strata, across the collapsings, or across the strata and the collapsings.

A possible test of homogeneity across the strata and the collapsings can be obtained by testing goodness–of–fit of the common proportional odds model (2.1), provided the within–stratum sample sizes are adequate. The test statistic is asymptotically chi–squared distributed with df= $cK-K-1$ when the model holds. A test of homogeneous odds ratios across the strata can be established by comparing the ML fit of the common proportional odds model (2.1) to a more complex proportional odds model with interaction terms between the explanatory variable Z and the control

variable W,

$$\log(\frac{\pi^*_{ijk}}{1-\pi^*_{ijk}}) = \alpha_{jk} + \theta_k I_{\{i=1\}} \ , \ j = 1, \ldots, c-1, \ k = 1, \ldots, K. \qquad (3.14)$$

The test is equal to a test of the null hypothesis $\theta_1 = \cdots = \theta_K$. A likelihood–ratio test statistic G^2 for the simpler (S) model (2.1), minus G^2 for the more complex (C) model (3.14), denoted by $G^2(\mathrm{S|C})$, can be used for testing that null hypothesis. When the simpler model S (2.1) holds, $G^2(\mathrm{S|C})$ has asymptotically a chi–squared distribution with df= $K-1$. The asymptotics again refer to adequate within–stratum sample sizes.

Another possible test of homogeneous odds ratios across the strata and the collapsings can be based on generalizing the Breslow–Day test. The Breslow–Day test statistic is a Pearson chi–squared test statistic, replacing the ML estimate with the ordinary Mantel–Haenszel estimate. We present a Pearson chi–squared type test statistic for testing a common cumulative odds ratio for the ordinal response case by using the common cumulative odds ratio estimate. The test statistic is directed at the potential lack of homogeneity of the odds ratios across both the strata and the collapsings, and it is not a model–based statistic. The deviations of observed and fitted values tend to be large when a particular odds ratio is much larger or much smaller than the common cumulative odds ratio estimate $\hat{\psi}$ (2.3). The generalized Breslow–Day test statistic Q^*_{BD} is as follows:

$$Q^*_{\mathrm{BD}} = \sum_{k=1}^{K} \sum_{j=1}^{c} \sum_{i=1}^{2} \frac{[X_{ijk} - \mathrm{E}(X_{ijk}|\hat{\psi})]^2}{\mathrm{E}(X_{ijk}|\hat{\psi})}, \qquad (3.15)$$

where $\mathrm{E}(X_{ijk}|\hat{\psi})$ is the expected value of X_{ijk} for $i \in \{1,2\}$, $j \in \{1, \ldots, c\}$, and $k \in \{1, \ldots, K\}$, under the common cumulative odds ratio assumption, treating $\hat{\psi}$ (2.3) as the true common cumulative odds ratio.

For each collapsed table, the expectation $\mathrm{E}(X^*_{1jk}|\hat{\psi})$ can be viewed as the expectation $\mathrm{E}(X_{11k}|\hat{\psi}_{\mathrm{MH}})$ in the Breslow–Day test statistic (3.12) for the binary response

case. Therefore, $\mathrm{E}(X^*_{1jk}|\hat{\psi})$ can be solved from a quadratic equation:

$$\hat{\psi} = \frac{\mathrm{E}(X^*_{1jk}|\hat{\psi})[n_{2k} - m^*_{jk} + \mathrm{E}(X^*_{1jk}|\hat{\psi})]}{[n_{1k} - \mathrm{E}(X^*_{1jk}|\hat{\psi})][m^*_{jk} - \mathrm{E}(X^*_{1jk}|\hat{\psi})]}, \tag{3.16}$$

where $m^*_{jk} = m_{1k} + \cdots + m_{jk}$ for $j \in \{1, \ldots, c-1\}$ and $k \in \{1, \ldots, K\}$. Since $\mathrm{E}(X_{1jk}|\hat{\psi})$ equals $\mathrm{E}(X^*_{1jk}|\hat{\psi}) - \mathrm{E}(X^*_{1(j-1)k}|\hat{\psi})$ and $\mathrm{E}(X_{2jk}|\hat{\psi})$ equals $m_{jk} - \mathrm{E}(X_{1jk}|\hat{\psi})$, the terms $\{\mathrm{E}(X_{ijk}|\hat{\psi})\}$ in expression (3.15) can be obtained by solving the equations (3.16) for all $c-1$ collapsings. One solution yields a possible value for a set of $\{\mathrm{E}(X_{1jk}|\hat{\psi})\}$, in the sense that expected values of $\{X_{ijk}\}$ are all positive.

For the test statistic Q^*_{BD}, if the common cumulative odds ratio estimate $\hat{\psi}$ is replaced by the ML estimate for the common proportional odds model (2.1), the test statistic is a goodness-of-fit statistic that has asymptotically a chi-squared distribution with df= $cK - K - 1$ when the model holds. Based on Tarone's arguments, we expect that Q^*_{BD} (3.15) using the Mantel-Haenszel-type estimator, an inefficient estimator, is stochastically larger under the null hypothesis of a common cumulative odds ratio than a chi-squared random variable with df= $cK - K - 1$.

All tests discussed above apply only for the large-stratum situation. However, under the sparse-data situation, we have not constructed a test of the common odds ratio assumption across both the strata and the collapsings. Next, we will propose a test of the hypothesis of a common odds ratio across the collapsings only. This test is good for both limiting situations. For $2 \times c \times K$ contingency tables, we collapse the response into a binary outcome $(\leq j, > j)$ for $j \in \{1, \ldots, c-1\}$. We assume that the odds ratios are constant across the strata within each collapsed $2 \times 2 \times K$ contingency table. Denote the common odds ratio for the jth collapsed table by ψ_j. We propose a test of equal common odds ratios across the $c-1$ collapsed tables. The null hypothesis for the test is H_0: $\psi_1 = \psi_2 = \cdots = \psi_{c-1}$ and the alternative hypothesis is H_a: at least one of $\{\psi_j\}$ differs from the others. Under the null hypothesis, the proportional odds assumption holds.

For the jth collapsed $2 \times 2 \times K$ contingency table, denote the ordinary Mantel–Haenszel estimator for the common odds ratio by $\hat{\psi}_j$,

$$\hat{\psi}_j = \frac{\sum_{k=1}^{K} R_{jk}}{\sum_{k=1}^{K} S_{jk}} = \frac{\sum_{k=1}^{K} X^*_{1jk}(n_{2k} - X^*_{2jk})/N_k}{\sum_{k=1}^{K} X^*_{2jk}(n_{1k} - X^*_{1jk})/N_k}.$$

Under the null hypothesis, we let all common odds ratios equal ψ. Similar to the arguments in Chapter 2, the ordinary Mantel–Haenszel estimators $\{\hat{\psi}_j, j = 1, \ldots, c-1\}$ follow asymptotically a multivariate normal distribution with mean $\boldsymbol{\psi} = (\psi, \ldots, \psi)'$ and asymptotic variance and covariance term $\boldsymbol{\Sigma}_\psi$ under the null hypothesis. Next, we construct a test statistic by using a quadratic form of the multivariate normal vector.

We propose a test treating $\hat{\psi}_1$ as a baseline and then compare each of $\hat{\psi}_j$ with the baseline. The test is invariant to the choice of the baseline. Intuitively, the property of invariance in this case is similar to the invariance of the testing of no treatment effects in the linear model case, where the baseline can be any one of the $c - 1$ treatment. Also, we use the log scale of the ordinary Mantel–Haenszel estimators to enhance convergence to normality. The asymptotic distribution for $\{\log \hat{\psi}_j\}$ follows by the delta method. A quadratic form proposed based on the multivariate normal distribution is

$$\boldsymbol{\Lambda}'(\boldsymbol{\Sigma}^*_\psi)^{-1}\boldsymbol{\Lambda}, \tag{3.17}$$

where

$$\boldsymbol{\Lambda}' = (\log \hat{\psi}_2 - \log \hat{\psi}_1, \ldots, \log \hat{\psi}_{c-1} - \log \hat{\psi}_1)$$

and the entries of $\boldsymbol{\Sigma}^*_\psi$ are based on the variances and covariances for $\{\log \hat{\psi}_j, j = 1, \ldots, c-1\}$. The quadratic form (3.17) has asymptotically a chi-squared distribution with $c - 2$ degrees of freedom under the null hypothesis. Since $\hat{\psi}$ is a consistent estimator for ψ, letting $\hat{\boldsymbol{\Sigma}}^*_\psi$ be a consistent estimator for $\boldsymbol{\Sigma}^*_\psi$, and applying Slutsky's theorem,

$$\boldsymbol{\Lambda}'(\hat{\boldsymbol{\Sigma}}^*_{\hat{\psi}})^{-1}\boldsymbol{\Lambda} \tag{3.18}$$

also has asymptotically a chi-squared distribution with df=$c-2$ under the null hypothesis.

Because the asymptotic property holds for the ordinary Mantel–Haenszel estimators under both limiting situations, we consider dually consistent estimators for the variances and covariances for $\{\log \hat{\psi}_j, j = 1, \ldots, c-1\}$. Under the null hypothesis, the dually consistent estimator for the variance of $\log \hat{\psi}_j$ follows from Chapter 2, which is

$$\hat{\mathrm{V}}\mathrm{ar}(\log \hat{\psi}_j) = C_j = \frac{\sum_{k=1}^{K} \hat{\phi}_{jk}(\hat{\psi})}{\hat{\psi}^2[\sum_{k=1}^{K} S_{jk}]^2},$$

where

$$\begin{aligned}\hat{\phi}_{jk}(\psi) &= \frac{n_{1k}n_{2k}}{N_k^2}\{\frac{\psi(n_{1k} - X^*_{1jk})X^*_{2jk}}{n_{1k}}[1 + (\psi - 1)\frac{X^*_{2jk}}{n_{2k}}] \\ &\quad + \frac{X^*_{1jk}(n_{2k} - X^*_{2jk})}{n_{2k}}[\psi - (\psi - 1)\frac{X^*_{1jk}}{n_{1k}}]\}.\end{aligned}$$

Also, the estimator for the covariance between $\log \hat{\psi}_j$ and $\log \hat{\psi}_{j'}$ for $j < j'$ is

$$\hat{\mathrm{C}}\mathrm{ov}(\log \hat{\psi}_j, \log \hat{\psi}_{j'}) = C_{jj'} = \frac{\sum_{k=1}^{K} \hat{\phi}_{jj'k}(\hat{\psi})}{\hat{\psi}^2[\sum_{k=1}^{K} S_{jk}][\sum_{k=1}^{K} S_{j'k}]},$$

where

$$\begin{aligned}\hat{\phi}_{jj'k}(\psi) &= \frac{n_{1k}n_{2k}}{N_k^2}\{\frac{\psi(n_{1k} - X^*_{1j'k})X^*_{2jk}}{n_{1k}}[1 + (\psi - 1)\frac{X^*_{2j'k}}{n_{2k}}] \\ &\quad + \frac{X^*_{1jk}(n_{2k} - X^*_{2j'k})}{n_{2k}}[\psi - (\psi - 1)\frac{X^*_{1j'k}}{n_{1k}}]\}.\end{aligned}$$

Thus, $\hat{\text{Var}}\ (\log\hat{\psi}_j - \log\hat{\psi}_1) = C_j + C_1 - 2C_{1j}$ and $\hat{\text{Cov}}\ (\log\hat{\psi}_j - \log\hat{\psi}_1,\ \log\hat{\psi}_{j'} - \log\hat{\psi}_1) = C_{jj'} - C_{1j} - C_{1j'} + C_1$ for all $j \neq j' \in \{2, \ldots, c-1\}$. Also, the symmetric matrix $\hat{\boldsymbol{\Sigma}}^*_{\hat{\psi}}$ in the quadratic form (3.18) can be written as

$$\hat{\boldsymbol{\Sigma}}^*_{\hat{\psi}} = \begin{bmatrix} C_2 + C_1 - 2C_{12} & C_{23} - C_{12} - C_{13} + C_1 & \cdots & C_{2(c-1)} - C_{12} - C_{1(c-1)} + C_1 \\ \vdots & & & \\ & C_{3(c-1)} - C_{1(c-1)} - C_{13} + C_1 & \cdots & C_{c-1} + C_1 - C_{1(c-1)} \end{bmatrix}.$$

For the example shown in Section 2.6, the log ordinary Mantel–Haenszel estimates $\{\log\hat{\psi}_j\}$ are 0.164, 0.159, and 0.161 for $j = 1$, 2, and 3, respectively. Applying the proposed test of homogeneous odds ratios across the collapsings, the test statistic (3.18) has value 0.005. The test statistic has asymptotically a chi–squared distribution with df=2 under the null hypothesis. Therefore, we conclude that the null hypothesis cannot be rejected for this example. Thus, the proportional odds assumption holds.

3.5 Weighted Average of the Stratum–Specific Cumulative Odds Ratio

When the common cumulative odds ratio assumption does not hold, the proposed cumulative odds ratio estimate (2.3) is a descriptive measure that summarizes association across the strata and the collapsings. Indeed, it is an estimator for a weighted average of the stratum–specific cumulative odds ratio. Under this situation, the proposed dually consistent variance estimator (2.10) discussed in Chapter 2 for the estimator is not appropriate, since the proofs of consistency in Chapter 2 fail when the odds ratios are heterogeneous. A reasonable variance estimator for this weighted average of the stratum–specific cumulative odds ratio estimator is only appropriate under the large–strata situation.

For the large–strata model, we derived a general variance estimator based on Guilbaud's (1983) arguments in Section 2.4. The general variance estimator given in Lemma 2.4.1 does not assume that the odds ratios are constant across the strata and

the collapsings. Therefore, it is an appropriate variance for the weighted average of the stratum-specific cumulative odds ratio estimator.

The estimate of the asymptotic variance can be obtained by replacing the population values by sample values,

$$\begin{aligned}\hat{\mathrm{V}}\mathrm{ar}(\hat{\psi}) &= \frac{1}{(\sum_{l=1}^{K}\hat{t}_l)^2}\{\sum_{k=1}^{K}\frac{\hat{t}_k^2}{n_{1k}}[\sum_{a=1}^{c}p_{1ak}(\frac{\hat{\psi}_k}{\hat{A}_k}\sum_{b=a+1}^{c}(b-a)p_{2bk}-\frac{\hat{\psi}}{\hat{B}_k}\sum_{b=1}^{a-1}(a-b)p_{2bk})^2] \\ &\quad + \sum_{k=1}^{K}\frac{\hat{t}_k^2}{n_{2k}}[\sum_{b=1}^{c}p_{2bk}(\frac{\hat{\psi}_k}{\hat{A}_k}\sum_{a=1}^{b-1}(b-a)p_{1ak}-\frac{\hat{\psi}}{\hat{B}_k}\sum_{a=b+1}^{c}(a-b)p_{1ak})^2]\} \\ &\quad - \sum_{k}(n_{1k}^{-1}+n_{2k}^{-1})[\hat{w}_k(\hat{\psi}_k-\hat{\psi})]^2,\end{aligned}$$

where $\hat{A}_k = \sum_{a<b}^{c}(b-a)p_{1ak}p_{2bk}$ $\hat{B}_k = \sum_{a>b}^{c}(a-b)p_{1ak}p_{2bk}$, $\hat{\psi}_k = [\sum_{j=1}^{c}p_{1jk}^{*}(1-p_{2jk}^{*})]/[\sum_{j=1}^{c-1}p_{2jk}^{*}(1-p_{1jk}^{*})]$, $\hat{t}_k = (n_{1k}^{-1}+n_{2k}^{-1})^{-1}[\sum_{j=1}^{c-1}(1-p_{1jk}^{*})p_{2jk}^{*}]$, and $\hat{w}_k = \hat{t}_k/\sum_k \hat{t}_k$. The expression $\hat{\mathrm{V}}\mathrm{ar}(\hat{\psi})$ is an appropriate variance estimator for the weighted average of the stratum-specific cumulative odds ratio estimator under the large-stratum situation.

CHAPTER 4
EXTENSIONS OF THE ORDINAL ODDS RATIO ESTIMATION

4.1 Introduction

For an ordinal response, we have already discussed the common cumulative odds ratio estimation for $2 \times c \times K$ contingency tables in Chapter 2. The proposed Mantel–Haenszel–type estimator has good asymptotic properties under both limiting situations. Chapter 3 presented evaluations of the proposed ordinal odds ratio estimator. The estimator has little loss of efficiency compared to the ML estimator under the large–stratum situation; but it tends to be much better than the ML estimator when the data are sparse. Now, in this chapter, we extend the ordinal odds ratio estimation to $r \times c \times K$ contingency tables, with the explanatory variable having either r nominal or r ordinal categories. In addition, we will discuss another ordinal odds ratio based on an adjacent–category logit model for K $2 \times c$ contingency tables.

The Mantel–Haenszel–type estimator for the common cumulative odds ratios for $2 \times c \times K$ contingency tables can be generalized when the explanatory or row variable Z has more than 2 levels. Suppose the explanatory variable has r nominal categories. We consider a proportional odds model assuming different cumulative odds ratios for different pairs of rows but without the row variable and the control variable interaction. For a given pair of rows, $r \times c \times K$ contingency tables can be viewed as $2 \times c \times K$ contingency tables. Assuming that the proportional odds model holds, the cumulative odds ratio for that pair of rows is constant across the strata. For different pairs of rows, the associations may be different; that is, the common cumulative odds ratios may vary across different pairs of rows based on the proportional odds model.

Again, the ML estimators may not be consistent under the sparse–data situation. For this reason, we consider dually consistent Mantel–Haenszel type estimators as alternative estimators.

For considering a nominal row variable, we will utilize some existing Mantel–Haenszel methods applying for the nominal response case. Mickey and Elashoff (1985) proposed *generalized Mantel–Haenszel partial estimators* to measure partial associations for $2 \times c \times K$ contingency tables in which the response variable has c nominal categories. Their estimators combine information from all possible pairs of columns. Furthermore, Greenland (1989) provided dually consistent variance and covariance estimators for these generalized Mantel–Haenszel partial estimators. In addition, for $2 \times c \times K$ contingency tables with c nominal response categories, Liang (1987) proposed alternative dually consistent estimators based on estimating functions. Sato (1991) provided variance and covariance estimators that are consistent under both limiting situations for the estimators produced from Liang's (1987) estimating functions. This chapter focuses on generalizing the Mickey–Elashoff and Greenland methods.

For the ordinal response case, this chapter generalizes the Mickey and Elashoff (1985) generalized Mantel–Haenszel partial estimators to estimating the common cumulative odds ratios, by combining information for all pairs of partial cumulative odds ratios on rows. Also, we propose dually consistent variance and covariance estimators for these generalized partial estimators using Greenland's (1989) procedures.

Suppose the explanatory or row variable has r ordinal categories instead of r nominal categories. One way to summarize the association from several $r \times c$ contingency tables is to assume that the common cumulative odds ratio holds for all pairs of adjacent rows. A proportional odds model in which the scores in rows are assigned with equal–spacing implies a common cumulative odds ratio. The asymptotic properties for the ML estimator may fail under the sparse–data situation. The dually consistent Mantel–Haenszel–type estimator and its dually consistent variance estimator for

the case with the r ordinal row categories can be derived by appropriately combining the generalized Mantel–Haenszel–type partial estimators and their variance and covariance estimators for the case with the r nominal row categories.

Finally, we propose an alternative ordinal odds ratio model to the cumulative odds ratio model for $2 \times c \times K$ contingency tables with c ordinal response categories. This model summarizes the association between the response variable and the explanatory variable conditional on strata using adjacent–category logits. The odds ratio is then constant across each adjacent pair of columns and across the strata. The unconditional ML estimator for the parameter in the adjacent–category logit model may not be consistent under the sparse–data situation. We propose an alternative Mantel–Haenszel–type estimator for it and a variance estimator by appropriately combining the generalized Mantel–Haenszel partial estimators and their variance and covariance estimators, for the case of $2 \times c \times K$ contingency tables with c nominal response categories.

Section 4.2 discusses ordinal–nominal cumulative odds ratios for $r \times c \times K$ contingency tables. Section 4.3 discusses ordinal–ordinal cumulative odds ratios for $r \times c \times K$ contingency tables. Section 4.4 presents estimators for alternative ordinal odds ratios for $2 \times c \times K$ contingency tables.

4.2 Ordinal–Nominal Cumulative Odds Ratios for $r \times c \times K$ Contingency Tables

In this section, we discuss ordinal–nominal odds ratios. The 'ordinal' means that the response has c ordinal categories. The 'nominal' represents an explanatory variable with r nominal categories. For the ordinal response, we again consider the proportional odds form of model, having a common cumulative odds ratio across the strata for each given pair of rows. In the model, the common cumulative odds ratios may vary for different pairs of rows. For all possible pairs of rows, $r \times c \times K$ contingency tables can be viewed as several $2 \times c \times K$ partial contingency tables. We use

the partial common cumulative odds ratios to summarize the associations between the response variable and the explanatory variable for those tables.

First, we illustrate the partial common cumulative odds ratios by fitting a proportional odds model without the explanatory variable and the control variable interaction. Consider a series of K $r \times c$ tables formed by r independent multinomial observations $X_{1jk}, \ldots, X_{rjk}$, $(j = 1, \ldots, c)$ with denominator $n_{1k}, \ldots, n_{rk}$ and probabilities $\pi_{1jk}, \ldots, \pi_{rjk}$, $(j = 1, \ldots, c)$ satisfying $\pi_{i1k} + \cdots + \pi_{ick} = 1$ for $i \in \{1, \ldots, r\}$ and $k \in \{1, \ldots, K\}$. When the explanatory variable Z has r nominal categories and the response variable Y has c ordinal categories, we consider the proportional odds model

$$\log\left(\frac{\pi^*_{ijk}}{1 - \pi^*_{ijk}}\right) = \alpha_{jk} + \Delta_i\,, \tag{4.1}$$

where $i \in \{1, \ldots, r\}$, $j \in \{1, \ldots, c-1\}$, and $k \in \{1, \ldots, K\}$. The terms $\{\Delta_i\}$ are called *row effects.*

From the proportional odds model (4.1), the odds ratio for the jth collapsed responses in the ith and the hth rows is

$$\frac{\pi^*_{ijk}(1 - \pi^*_{hjk})}{(1 - \pi^*_{ijk})\pi^*_{hjk}} = \exp(\Delta_i - \Delta_h) = \psi_{ih}\,, \tag{4.2}$$

for $j \in \{1, \ldots, c-1\}$ and $k \in \{1, \ldots, K\}$. We call ψ_{ih} the *partial common cumulative odds ratio* for rows i and h. Its interpretation is that the odds of making response $\leq j$ are $\exp(\Delta_i - \Delta_h)$ times higher at level i of the explanatory variable Z than at level h of Z, for each stratum.

Estimates for the partial common cumulative odds ratio can be obtained from ML fitting of the proportional odds model (4.1). However, the ML estimators may not be consistent under the sparse–data situation. Alternatively, we can summarize the associations between the explanatory variable and the response variable from several $r \times c$ tables by using the Mantel–Haenszel method. When we consider rows i and h

only, we have a $2 \times c \times K$ partial contingency table. Thus, the problem reduces to estimating the common cumulative odds ratio in a $2 \times c \times K$ contingency table. A possible Mantel–Haenszel–type estimator for ψ_{ih} is simply the common cumulative odds ratio estimator discussed in Chapter 2 for the $2 \times c \times K$ table with rows i and h alone. We have already shown that the estimator is consistent under both sparse–data and large–stratum situations. Also, the dually consistent variance and covariance estimators for these Mantel–Haenszel–type partial estimators can be derived using the same arguments as in Chapter 2 under the partial common cumulative odds ratio assumption (4.2).

There is a disadvantage in constructing the Mantel–Haenszel–type partial estimator by considering a partial $2 \times c \times K$ contingency table with two particular rows alone. Begg and Gray (1984) discussed a related problem. Begg and Gray (1984) studied the performance of an individualized logistic regression, in which a series of separate simple logistic regression analyses are performed as a replacement for a polychotomous logistic regression. In their study, they wished to discriminate among r diagnostic categories and a baseline category. The polychotomous model is a logistic regression model based on all possible pairs of diagnostic categories against the baseline category. The individualized logistic regression considers simple logistic regressions for all diagnostic categories separately against the baseline category. They claimed that the use of individualized logistic regression has some loss of efficiency compared to the polychotomous logistic regression. Based on their findings, we may have some loss of efficiency for the Mantel–Haenszel–type partial estimator by considering a partial $2 \times c \times K$ contingency table with two rows alone, compared to an estimator using the full $r \times c \times K$ contingency table.

We propose a generalized Mantel–Haenszel–type partial estimator for ψ_{ih} that improves on the Mantel–Haenszel–type partial estimator that uses only one pair of

rows. The improved estimator combines all information for all partial $2 \times c \times K$ contingency tables. The arguments follow results of Mickey and Elashoff (1985).

Mickey and Elashoff (1985) proposed a generalized Mantel–Haenszel partial estimator for $2 \times c \times K$ contingency tables with c nominal response categories. For a nominal response case, a partial odds ratio estimator can also be obtained by considering two particular columns alone. For two particular columns, we have a $2 \times 2 \times K$ contingency table that is a part of the original $2 \times c \times K$ contingency table. A possible Mantel–Haenszel partial estimator is the ordinary Mantel–Haenszel estimator considering the $2 \times 2 \times K$ contingency table consisting of these two columns only. The Mickey–Elashoff generalized Mantel–Haenszel estimator combines all information from all these Mantel–Haenszel partial estimators. We define notations for the Mickey–Elashoff estimator next.

Consider a series of K $2 \times c$ contingency tables formed by two independent multinomial observations X_{1jk}, X_{2jk}, $(j = 1, \ldots, c)$ with denominators n_{1k}, n_{2k} and probabilities π_{1jk}, π_{2jk}, $(j = 1, \ldots, c)$ satisfying $\pi_{i1k} + \pi_{i2k} + \cdots + \pi_{ick} = 1$ for $i \in \{1, 2\}$ and $k \in \{1, \ldots, K\}$. Suppose the response has c nominal categories. Define $f_{jj'k} = X_{1jk}X_{2j'k}/N_k$ and $F_{jj'} = \sum_k f_{jj'k}$. Let $L^*_{jj'} = \log(F_{jj'}/F_{j'j})$. Then, $\exp(L^*_{jj'})$ is the ordinary Mantel–Haenszel estimator for K 2×2 tables, considering columns j and j' only and using the weight $1/N_k$.

Mickey and Elashoff (1985) noted that one could obtain an alternative estimator. The motivation of the estimator is from a loglinear model with no three–way interaction. Denote the cell mean for the ith and the jth column in the kth stratum by μ_{ijk}. Define the loglinear model as follows:

$$\log \mu_{ijk} = \lambda_0 + \lambda^Z_i + \lambda^Y_j + \lambda^W_k + \lambda^{ZY}_{ij} + \lambda^{ZW}_{ik} + \lambda^{YW}_{jk},$$

The loglinear model has constraints $\sum_i \lambda^Z_i = \sum_j \lambda^Y_j = \cdots = \sum_j \lambda^{YW}_{jk} = \sum_k \lambda^{YW}_{jk} = 0$. The relationship between the log odds ratios and the loglinear partial association

parameters (Bishop *et al.*, 1975, p. 26) is

$$\lambda_{1j}^{ZY} = \sum_{j'} (\log \psi_{jj'}^{\star})/(2c)\,, \tag{4.3}$$

where $\psi_{jj'}^{\star} = \pi_{1jk}\pi_{2j'k}/[\pi_{1j'k}\pi_{2jk}]$. The ordinary Mantel–Haenszel estimator $\exp(L_{jj'}^{\star})$ can be used to estimate $\psi_{jj'}^{\star}$. Therefore, using the equation (4.3), the generalized Mantel–Haenszel partial estimate for $\psi_{jj'}^{\star}$, based on estimates of λ_{1j}^{ZY} and $\lambda_{1j'}^{ZY}$, is defined by

$$\begin{aligned} \log(\hat{\psi}_{jj'}^{\star}) &= \bar{L}_{jj'}^{\star} = \hat{\lambda}_{1j}^{ZY} - \hat{\lambda}_{1j'}^{ZY} + \hat{\lambda}_{2j'}^{ZY} - \hat{\lambda}_{2j}^{ZY} = 2(\hat{\lambda}_{1j}^{ZY} - \hat{\lambda}_{1j'}^{ZY}) \\ &= (\sum_{j'} \log \hat{\psi}_{jj'}^{\star} - \sum_{j} \log \hat{\psi}_{j'j}^{\star})/c = (L_{j+}^{\star} - L_{j'+}^{\star})/c\,, \end{aligned} \tag{4.4}$$

where a subscript "+" denotes summation over the columns.

Since $L_{jj'}^{\star}$ is a dually consistent estimator of $\log(\psi_{jj'}^{\star})$, $\bar{L}_{jj'}^{\star}$ is a dually consistent estimator of $\log(\psi_{jj'}^{\star})$ as well. The arguments of consistency and asymptotically normal distribution are similar to the arguments in Chapter 2. Dually consistent estimators of variances and covariances of $\{\bar{L}_{jj'}^{\star}\}$ were proposed by Greenland (1989). The partial odds ratios are mutually compatible, in the sense that $\psi_{jl}^{\star}\psi_{lj'}^{\star} = \psi_{jj'}^{\star}$. The estimators $\{\bar{L}_{jj'}^{\star}\}$ are preferable to $\{L_{jj'}^{\star}\}$ because they are also mutually compatible, in the sense that $\bar{L}_{jl}^{\star} + \bar{L}_{lj'}^{\star} = \bar{L}_{jj'}^{\star}$ (*i. e.*, $\hat{\psi}_{jl}^{\star}\hat{\psi}_{lj'}^{\star} = \hat{\psi}_{jj'}^{\star}$). In addition, $\bar{L}_{jj'}^{\star}$ can have worthwhile efficiency advantages over $L_{jj'}^{\star}$ (Greenland, 1989).

For the case of an ordinal response and nominal explanatory variable, we generalize the arguments to estimate the partial common cumulative odds ratio, ψ_{ih}, for rows i and h. Define $R_{jk}^{ih} = X_{ijk}^{\star}(n_{hk} - X_{hjk}^{\star})/N_k$ and $S_{jk}^{ih} = X_{hjk}^{\star}(n_{ik} - X_{ijk}^{\star})/N_k$, for all $i \neq h \in \{1,\ldots,r\}$, $j \in \{1,\ldots,c-1\}$, and $k \in \{1,\ldots,K\}$. Let $L_{ih} = \log(\sum_k \sum_j R_{jk}^{ih}/\sum_k \sum_j S_{jk}^{ih})$. The estimator $\exp(L_{ih})$ is the Mantel–Haenszel-type estimator for the common cumulative odds ratio discussed in Chapter 2, with weight $1/N_k$, considering rows i and h only. We call $\exp(L_{ih})$ a *Mantel–Haenszel-type partial*

estimator. Notice that for two particular rows, the weights of the terms R_{jk}^{ih} and S_{jk}^{ih} should be $1/(n_{ik}+n_{hk})$ instead of $1/N_k$ by the definition of the common cumulative odds ratio estimator proposed in Chapter 2. For $2\times 2\times K$ contingency tables, as Mantel and Haenszel (1959) noted, under the null hypothesis and using the large–stratum model, the weight $1/(n_{ik}+n_{hk})$ yields an inverse–null–variance weighted odds ratio estimator. So, L_{ih} with the weight $1/N_k$ is expected to be less efficient than the one with the weight $1/(n_{ik}+n_{hk})$. Regardless of the weights, they are both dually consistent for $\log(\psi_{ih})$ and have asymptotically normal distributions by the arguments given in Chapter 2.

Following the Mickey–Elashoff approach, we propose the *generalized Mantel–Haenszel–type partial estimator* for ψ_{ih}, defined by

$$\log(\hat{\psi}_{ih}) = \bar{L}_{ih} = (L_{i+} - L_{h+})/r\,,$$

where a subscript "+" denotes summation over the rows. This measures the association between the explanatory variable and the response variable for particular rows i and h, by considering all possible pairs of associations. Since L_{ih} is dually consistent and asymptotically normally distributed, $\bar{L}_{ih}$ is dually consistent and asymptotically normally distributed as well. Similarly, $\bar{L}_{ih}$ is mutually compatible in the sense that $\bar{L}_{ih}+\bar{L}_{hi'} = \bar{L}_{ii'}$ (*i. e.*, $\hat{\psi}_{ih}\hat{\psi}_{hi'} = \hat{\psi}_{ii'}$). For $2\times c\times K$ contingency tables with c nominal response categories, Greenland (1989) noted that the weighted estimators are mutually incompatible (*i. e.*, $\bar{L}_{jl}^{*}+\bar{L}_{lj'}^{*} \neq \bar{L}_{jj'}^{*}$) unless the weights do not vary, and so on this basis the simple average estimators with equal weights ($1/N_k$) appear preferable.

We follow the same procedures as in Chapter 2 to obtain dually consistent variance and covariance estimators for $\{L_{ih}\}$. Since $\bar{L}_{ih}$ is a linear combination of $\{L_{ih}, i\neq h\in\{1,\ldots r\}\}$, estimators for $\mathrm{Var}(\bar{L}_{ih})$ and $\mathrm{Cov}(\bar{L}_{ih},\ \bar{L}_{i'h'})$ follow immediately from variances and covariances for $\{L_{ih}, i\neq h\in\{1,\ldots r\}\}$. Here, we consider estimators

with the weights $1/N_k$ only, since these estimators have simpler expressions. For the estimators with the weights $1/(n_{ik}+n_{hk})$, results can be easily modified by replacing N_k with $(n_{ik}+n_{hk})$. Let U_{ihg} represent the covariance estimator for L_{ih} and L_{ig}. Consequently, the U_{ihh} is the variance estimator for L_{ih}. Chapter 2 presented a dually consistent estimator of $\mathrm{Var}(\hat{\psi}_{ih})$. Using the delta method, the variance of L_{ih} for all $i \neq h$ is as follows:

$$U_{ihh} = \frac{1}{(\hat{\psi}_{ih})^2} \frac{\sum_{k=1}^{K} \hat{\xi}_k^{ih}(\hat{\psi}_{ih})}{[\sum_{k=1}^{K} \sum_{j=1}^{c-1} S_{jk}^{ih}]^2},$$

where $\hat{\psi}_{ih} = \sum_k \sum_j R_{jk}^{ih} / \sum_k \sum_j S_{jk}^{ih}$, $\hat{\xi}_k^{ih}(\psi) = \sum_{j=1}^{c-1} \hat{\phi}_{jk}^{ih}(\psi) + 2\sum_{j<j'=1}^{c-1} \hat{\phi}_{jj'k}^{ih}(\psi)$,

$$\begin{aligned}\hat{\phi}_{jk}^{ih}(\psi) &= \frac{n_{ik}n_{hk}}{N_k^2}\{\frac{\psi(n_{ik}-X_{ijk}^{*})X_{hjk}^{*}}{n_{ik}}[1+(\psi-1)\frac{X_{hjk}^{*}}{n_{hk}}] \\ &\quad + \frac{X_{ijk}^{*}(n_{hk}-X_{hjk}^{*})}{n_{hk}}[\psi-(\psi-1)\frac{X_{ijk}^{*}}{n_{ik}}]\}\end{aligned}$$

$$\begin{aligned}\hat{\phi}_{jj'k}^{ih}(\psi) &= \frac{n_{ik}n_{hk}}{N_k^2}\{\frac{\psi(n_{ik}-X_{ij'k}^{*})X_{hjk}^{*}}{n_{ik}}[1+(\psi-1)\frac{X_{hj'k}^{*}}{n_{hk}}] \\ &\quad + \frac{X_{ijk}^{*}(n_{hk}-X_{hj'k}^{*})}{n_{hk}}[\psi-(\psi-1)\frac{X_{ij'k}^{*}}{n_{ik}}]\}.\end{aligned}$$

The variance of L_{ih} and L_{hi} is equivalent, since $L_{ih} = -L_{hi}$; that is, the variance of L_{ih} is invariant under the interchange of rows i and h for $r \times c \times K$ contingency tables. Our proposed dually consistent variance estimator, U_{ihh}, is invariant when we interchange the rows. It is not difficult to show the equivalence, since

$$U_{hii} = (\hat{\psi}_{ih})^2 \frac{\sum_{k=1}^{K} \hat{\xi}_k^{hi}(\hat{\psi}_{hi})}{[\sum_{k=1}^{K} \sum_{j=1}^{c-1} R_{jk}^{ih}]^2} = \frac{\sum_{k=1}^{K} \hat{\xi}_k^{hi}(\hat{\psi}_{hi})}{[\sum_{k=1}^{K} \sum_{j=1}^{c-1} S_{jk}^{ih}]^2},$$

and the first term of the numerator part of U_{hii} simplifies to

$$\begin{aligned}
\hat{\phi}^{hi}_{jk}(\hat{\psi}_{hi}) &= \frac{n_{ik}n_{hk}}{N_k^2}\{\frac{(n_{hk}-X^*_{hjk})X^*_{ijk}}{\hat{\psi}_{ih}n_{ik}}[1+(\frac{1}{\hat{\psi}_{ih}}-1)\frac{X^*_{ijk}}{n_{ik}}] \\
&\quad +\frac{X^*_{hjk}(n_{ik}-X^*_{ijk})}{n_{ik}}[\frac{1}{\hat{\psi}_{ih}}-(\frac{1}{\hat{\psi}_{ih}}-1)\frac{X^*_{hjk}}{n_{hk}}]\} \\
&= \frac{n_{ik}n_{hk}}{\hat{\psi}^2_{ih}N_k^2}\{\frac{\hat{\psi}_{ih}(n_{ik}-X^*_{ijk})X^*_{hjk}}{n_{ik}}[1+(\hat{\psi}_{ih}-1)\frac{X^*_{hjk}}{n_{hk}}] \\
&\quad +\frac{X^*_{ijk}(n_{hk}-X^*_{hjk})}{n_{hk}}[\hat{\psi}_{ih}-(\hat{\psi}_{ih}-1)\frac{X^*_{ijk}}{n_{ik}}]\}=\frac{1}{(\hat{\psi}_{ih})^2}\hat{\phi}^{ih}_{jk}(\hat{\psi}_{ih}).
\end{aligned}$$

Similarly, the second term simplifies to $\hat{\phi}^{hi}_{jj'k}(\hat{\psi}_{hi}) = \hat{\phi}^{ih}_{jj'k}(\hat{\psi}_{ih})/(\hat{\psi}_{ih})^2$. Therefore, $U_{ihh} = U_{hii}$.

The covariance of L_{ih} and L_{ig} can be obtained using similar arguments to those in Chapter 2. Also, the covariance is invariant under the interchange of the rows, in the sense that Cov(L_{ih}, L_{ig})=Cov(L_{ig}, L_{ih})= −Cov(L_{gi}, L_{ih})= −Cov(L_{hi}, L_{ig})=Cov(L_{gi}, L_{hi})=Cov(L_{hi}, L_{gi}). Since our proposed dually consistent covariance estimator is not invariant when we interchange the rows, we propose a dually consistent estimator U_{ihg} that is the arithmetic average of them. Then, the dually consistent estimator of Cov(L_{ih}, L_{ig}) for all $i \neq h \neq g$, is given by

$$\begin{aligned}
U_{ihg} &= \frac{1}{6}[\hat{\mathrm{C}}\mathrm{ov}(L_{ih},L_{ig})+\hat{\mathrm{C}}\mathrm{ov}(L_{ig},L_{ih})-\hat{\mathrm{C}}\mathrm{ov}(L_{gi},L_{ih})-\hat{\mathrm{C}}\mathrm{ov}(L_{hi},L_{ig}) \\
&\quad +\hat{\mathrm{C}}\mathrm{ov}(L_{gi},L_{hi})+\hat{\mathrm{C}}\mathrm{ov}(L_{hi},L_{gi})].
\end{aligned}$$

The term $\hat{\mathrm{C}}\mathrm{ov}(L_{ih}, L_{ig})$ measures the covariance with the row i on the top, h in the middle, and g on the bottom; the term $\hat{\mathrm{C}}\mathrm{ov}(L_{gi}, L_{ih})$ measures the covariance with the row g on the top, i in the middle, and h on the bottom; and the term $\hat{\mathrm{C}}\mathrm{ov}(L_{gi}, L_{hi})$ measures the covariance with the row g on the top, h in the middle, and i on the bottom. Let a and b represent either h or g. Each covariance term in U_{ihg} is as

follows:

$$\begin{aligned}
\hat{\mathrm{Cov}}(L_{ia}, L_{ib}) &= \frac{1}{\hat{\psi}_{ia}\hat{\psi}_{ib}}\hat{\mathrm{Cov}}(\hat{\psi}_{ia}, \hat{\psi}_{ib}) \quad \text{, for all } i \neq a \neq b \\
&= \frac{1}{\hat{\psi}_{ia}\hat{\psi}_{ib}}[\frac{\sum_{k=1}^{K}\sum_{j=1}^{c-1}\hat{\mathrm{Cov}}(R_{jk}^{ia} - \psi_{ia}S_{jk}^{ia}, R_{jk}^{ib} - \psi_{ib}S_{jk}^{ib})}{(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ia})(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ib})} \\
&+ \frac{\sum_k \sum_{j'\neq j}^{c-1}\hat{\mathrm{Cov}}(R_{jk}^{ia} - \psi_{ia}S_{jk}^{ia}, R_{j'k}^{ib} - \psi_{ib}S_{j'k}^{ib})}{(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ia})(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ib})}] \\
&= \frac{1}{\hat{\psi}_{ia}\hat{\psi}_{ib}}\frac{\sum_k \sum_{j=1}^{c-1}\hat{\phi}_{jk}^{iaib}(\hat{\psi}_{ia}, \hat{\psi}_{ib}) + \sum_k \sum_{j\neq j'}^{c-1}\hat{\phi}_{jj'k}^{iaib}(\hat{\psi}_{ia}, \hat{\psi}_{ib})}{(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ia})(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ib})},
\end{aligned}$$

$$\hat{\phi}_{jk}^{iaib}(\psi_{ia}, \psi_{ib}) = \frac{n_{ik}n_{ak}n_{bk}}{N_k^2}[\frac{X_{ijk}^* X_{ajk}^*}{n_{ik}n_{ak}}(\psi_{ia} - 1) + \frac{X_{ijk}^*}{n_{ik}} - \psi_{ia}\frac{X_{ajk}^* X_{bjk}^*}{n_{ak}n_{bk}}]$$

$$\begin{aligned}
\hat{\phi}_{jj'k}^{iaib}(\psi_{ia}, \psi_{ib}) &= \frac{n_{ik}\psi_{ia}}{N_k^2}[X_{ajk}^*(n_{bk} - X_{bj'k}^*)] \text{ if } j < j' \\
&= \frac{n_{ik}\psi_{ib}}{N_k^2}[X_{bj'k}^*(n_{ak} - X_{ajk}^*)] \text{ if } j > j',
\end{aligned}$$

$$\begin{aligned}
\hat{\mathrm{Cov}}(L_{ai}, L_{ib}) &= \frac{1}{\hat{\psi}_{ai}\hat{\psi}_{ib}}\hat{\mathrm{Cov}}(\hat{\psi}_{ai}, \hat{\psi}_{ib}) \quad \text{, for all } i \neq a \neq b \\
&= \frac{1}{\hat{\psi}_{ai}\hat{\psi}_{ib}}[\frac{\sum_{k=1}^{K}\sum_{j=1}^{c-1}\hat{\mathrm{Cov}}(R_{jk}^{ai} - \psi_{ai}S_{jk}^{ai}, R_{jk}^{ib} - \psi_{ib}S_{jk}^{ib})}{(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ai})(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ib})} \\
&+ \frac{\sum_k \sum_{j'\neq j}^{c-1}\hat{\mathrm{Cov}}(R_{jk}^{ai} - \psi_{ai}S_{jk}^{ai}, R_{j'k}^{ib} - \psi_{ib}S_{j'k}^{ib})}{(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ai})(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ib})}] \\
&= \frac{1}{\hat{\psi}_{ai}\hat{\psi}_{ib}}\frac{\sum_k \sum_{j=1}^{c-1}\hat{\phi}_{jk}^{aiib}(\hat{\psi}_{ai}, \hat{\psi}_{ib}) + \sum_k \sum_{j\neq j'}^{c-1}\hat{\phi}_{jj'k}^{aiib}(\hat{\psi}_{ai}, \hat{\psi}_{ib})}{(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ai})(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ib})},
\end{aligned}$$

$$\hat{\phi}_{jk}^{aiib}(\psi_{ai}, \psi_{ib}) = \frac{n_{ik}n_{bk}n_{ak}}{N_k^2}[(\psi_{ai}-1)(\frac{X_{ajk}^* X_{ijk}^*}{n_{ak}n_{ik}} - \frac{X_{ajk}^* X_{ijk}^* X_{bjk}^*}{n_{ik}n_{ak}n_{bk}}) - \psi_{ai}\frac{X_{ijk}^*}{n_{ik}} + \psi_{ai}\frac{X_{ijk}^* X_{bjk}^*}{n_{ik}n_{bk}}],$$

$$\hat{\phi}^{aiib}_{jj'k}(\psi_{ai},\psi_{ib}) = \frac{n_{ik}}{N_k^2}[-X^*_{ajk}(n_{bk}-X^*_{bj'k})] \text{ if } j<j'$$

$$= \frac{n_{ik}\psi_{ai}\psi_{ib}}{N_k^2}[-(n_{ak}-X^*_{ajk})X^*_{bj'k}] \text{ if } j>j',$$

$$\hat{\mathrm{C}}\mathrm{ov}(L_{ai},L_{bi}) = \frac{\sum_k\sum_{j=1}^{c-1}\hat{\phi}^{aibi}_{jk}(\hat{\psi}_{ai},\hat{\psi}_{bi}) + \sum_k\sum_{j\neq j'}^{c-1}\hat{\phi}^{aibi}_{jj'k}(\hat{\psi}_{ai},\hat{\psi}_{bi})}{(\hat{\psi}_{ai}\hat{\psi}_{bi})(\sum_k\sum_{j=1}^{c-1}S^{bi}_{jk})(\sum_k\sum_{j=1}^{c-1}S^{ai}_{jk})}, \text{for all } i\neq a\neq b$$

$$\hat{\phi}^{aibi}_{jk}(\psi_{ai},\psi_{bi}) = \frac{n_{ik}n_{ak}n_{bk}}{N_k^2}[\psi_{ai}(\frac{-X^*_{ajk}X^*_{bjk}}{n_{ak}n_{bk}}) + \psi_{ai}\frac{X^*_{bjk}}{n_{bk}}],$$

$$\hat{\phi}^{aibi}_{jj'k}(\psi_{ai},\psi_{bi}) = \frac{n_{ik}\psi_{bi}}{N_k^2}[X^*_{ajk}(n_{bk}-X^*_{bj'k})] \text{ if } j<j'$$

$$= \frac{n_{ik}\psi_{ai}}{N_k^2}[X^*_{bj'k}(n_{ak}-X^*_{ajk})] \text{ if } j>j'.$$

Both U_{ihh} and U_{ihg} are invariant under the interchange of the rows. Also, under reversing ordering of the response, in the sense that the responses $1,2,\ldots,c$ are reversed to $c,c-1,\ldots,1$, the new L_{ih} based on the reverse–ordered response equals $-L_{ih}$ based on the original response. Therefore, the variances and covariances of $\{L_{ih}\}$ are invariant under reversing ordering of the response; that is, they are palindromic invariant. The estimators U_{ihh} and U_{ihg} are not palindromic invariant. We may use an average of the estimates from the original ordering of the response and the reverse–ordered response to obtain a symmetric version of the variance and covariance estimates. However, when the partial common cumulative odds ratios assumption holds, there is little difference between the symmetric version of the estimates and the non–symmetric version of the estimates. The Pitman ARE and an example shown later are based on the symmetric version of the variance and covariance estimates.

Dually consistent variance and covariance estimators for $\{\bar{L}_{ih}\}$ follow from basic formulas for variances and covariances of sums. Greenland (1989) showed these

formulas for $\mathrm{Var}(\bar{L}_{ih})$ and $\mathrm{Cov}(\bar{L}_{ih}, \bar{L}_{i'h'})$, based on variances and covariances of $\{L_{ih}\}$. Define $U_{ihg} = 0$ if $i = h$ or g, $U^+_{ih} = U_{i++}$ if $i = h$, and $U^+_{ih} = U^+_{hi} = U_{+ih} - U_{ih+} - U_{hi+} + U_{ihh}$ if $i \neq h$. Since

$$\begin{aligned}
\mathrm{Cov}(L_{i+}, L_{h+}) &= \mathrm{Cov}\Big(\sum_{i' \neq i=1}^{r} L_{ii'}, \sum_{h' \neq h=1}^{r} L_{hh'}\Big) \\
&= \mathrm{Cov}\Big(\sum_{i' \neq i=1}^{r} L_{ii'}, L_{hi}\Big) + \mathrm{Cov}\Big(L_{ih}, \sum_{h' \neq h=1}^{r} L_{hh'}\Big) \\
&\quad + \sum_{i' \neq (i \cup h)=1}^{r} \mathrm{Cov}(L_{ii'}, L_{hi'}) - \mathrm{Cov}(L_{ih}, L_{hi}) \\
&= -\mathrm{Cov}\Big(\sum_{i' \neq i=1}^{r} L_{ii'}, L_{ih}\Big) - \mathrm{Cov}\Big(L_{hi}, \sum_{h' \neq h=1}^{r} L_{hh'}\Big) \\
&\quad + \sum_{i' \neq (i \cup h)=1}^{r} \mathrm{Cov}(L_{i'i}, L_{i'h}) + \mathrm{Cov}(L_{ih}, L_{ih}),
\end{aligned}$$

the dually consistent estimator for $\mathrm{Cov}(L_{i+}, L_{h+})$ is U^+_{ih}. If $i = h$, the covariance of L_{i+} and L_{h+} is $\sum_{i' \neq (i \cup h)=1}^{r} \mathrm{Cov}(L_{ii'}, L_{hi'}) = \sum_{i' \neq i=1}^{r} \mathrm{Cov}(L_{ii'}, L_{ii'})$. The dually consistent estimator for it is U_{i++}. Then, a dually consistent estimator for the covariance of $\bar{L}_{ih}$, $\bar{L}_{i'h'}$ is

$$\hat{\mathrm{C}}\mathrm{ov}(\bar{L}_{ih}, \bar{L}_{i'h'}) = \frac{1}{r^2} \hat{\mathrm{C}}\mathrm{ov}(L_{i+} - L_{h+}, L_{i'+} - L_{h'+}) = \frac{1}{r^2}(U^+_{ii'} - U^+_{ih'} - U^+_{hi'} + U^+_{hh'}).$$

If $i = i'$ and $h = h'$, this formula yields a dually consistent estimator for $\mathrm{Var}(\bar{L}_{ih})$:

$$\frac{1}{r^2}(U_{i++} - 2U^+_{ih} + U_{h++}).$$

Now, we study the Pitman ARE for L_{ih} compared to $\bar{L}_{ih}$ under the large–stratum situation. We expect that L_{ih} has some loss of efficiency compared to the improved estimator $\bar{L}_{ih}$, since L_{ih} uses information only from a single $2 \times c \times K$ contingency table. Begg and Gray (1984) discussed a related study to compare the performance

of the individualized logistic regression and the performance of the polychotomous logistic regression. They claimed that the Pitman ARE for the estimators comparing two models was 1 when there is no difference for all r diagnostic categories and the baseline category; that is, when all parameters equal 0. We will show that our study has similar results.

Suppose there is only one stratum. Under the null hypothesis of $\psi_{ih} = 1$ for all i and h, we now show that the Pitman asymptotic relative efficiency for L_{ih} compared to $\bar{L}_{ih}$ is 1. Since L_{ih} and $\bar{L}_{ih}$ have asymptotically normal distributions with the same mean, $\log \psi_{ih}$, the Pitman ARE can be obtained by comparing the two asymptotic variances. Usually, Pitman ARE is used to compare two test statistics. In our case, the two test statistics are point estimators. Therefore, the Pitman ARE is equivalent to the asymptotic relative efficiency for the two estimators. Let V^0_{ihh} denote the asymptotic variance of $\sqrt{N_+}L_{ih}$ and V^0_{ihg} denote the asymptotic covariance for $\sqrt{N_+}L_{ih}$ and $\sqrt{N_+}L_{ig}$ under the null hypothesis. The term V^0_{ihh} can simplify to

$$\begin{aligned}\mathrm{Var}(\sqrt{N_+}L_{ih}) &= V^0_{ihh} = (\frac{1}{a_i} + \frac{1}{a_h})[\frac{\sum_{j=1}^{c-1}(1-\pi^{0*}_{ij})\pi^{0*}_{ij} + 2\sum_{j<j'}^{c-1}(1-\pi^{0*}_{ij})\pi^{0*}_{ij'}}{(\sum_{j=1}^{c}\pi^{0*}_{ij}(1-\pi^{0*}_{ij}))^2}] \\ &= (\frac{1}{a_i} + \frac{1}{a_h})\Gamma,\end{aligned}$$

where $N_+a_i = n_i$, and $\Gamma = [\sum_{j=1}^{c-1}(1-\pi^{0*}_{ij})\pi^{0*}_{ij} + 2\sum_{j<j'}^{c-1}(1-\pi^{0*}_{ij})\pi^{0*}_{ij'}]/(\sum_{j=1}^{c}\pi^{0*}_{ij}(1-\pi^{0*}_{ij}))^2$. The asymptotic covariances under the null hypothesis are

$$V^0_{ihg} = \frac{1}{a_i}\Gamma = V^0_{igh}\,,\ V^0_{hig} = \frac{1}{a_h}\Gamma = V^0_{hgi}\,,\ V^0_{gih} = \frac{1}{a_g}\Gamma = V^0_{ghi}\,.$$

Also, the asymptotic variance for $\sqrt{N_+}\bar{L}_{ih}$ is

$$\begin{aligned}
\mathrm{Var}(\sqrt{N_+}\bar{L}_{ih}) &= \frac{1}{r^2}[V^0_{i++} - 2V^0_{ih+} + V^0_{h++}] \\
&= \frac{1}{r^2}\{\Gamma\sum_{j\neq i}^{r}(\frac{1}{a_i}+\frac{1}{a_j}) + 2\Gamma\sum_{j<j'\neq i}^{r}(\frac{1}{a_i}) + \Gamma\sum_{j\neq h}^{r}(\frac{1}{a_h}+\frac{1}{a_j}) + 2\Gamma\sum_{j<j'\neq h}^{r}(\frac{1}{a_h}) \\
&- 2\Gamma[\sum_{j\neq i,j\neq h}^{r}(\frac{1}{a_j}) - \sum_{j\neq i,j\neq h}^{r}(\frac{1}{a_i}) - (\frac{1}{a_i}+\frac{1}{a_h}) - \sum_{j\neq i,j\neq h}^{r}(\frac{1}{a_h})]\} \\
&= \frac{\Gamma}{r^2}[(r-1)\frac{1}{a_i} + \frac{1}{a_h} + (r-1)(r-2)\frac{1}{a_i} + (r-1)\frac{1}{a_h} + \frac{1}{a_i} \\
&+ (r-1)(r-2)\frac{1}{a_h} + 2(r-2)\frac{1}{a_i} + 2(\frac{1}{a_i}+\frac{1}{a_h}) + 2(r-2)\frac{1}{a_h}] \\
&= \Gamma(\frac{1}{a_i}+\frac{1}{a_h}) = \mathrm{Var}(\sqrt{N_+}L_{ih})\,.
\end{aligned}$$

Therefore, for the one–stratum situation, there is no efficiency loss for L_{ih} compared to $\bar{L}_{ih}$ under the null hypothesis.

Tables 4.1 to 4.4 show the asymptotic relative efficiency for L_{ih} and $\bar{L}_{ih}$ with different weights under different situations. In order to simplify our study, we let the response variable have three categories (*i. e.*, $c = 3$) and the explanatory variable have three or five categories (*i. e.*, $r = 3$ or 5). Also, we let the number of strata equal one or three (*i. e.*, $K = 1$ or 3), with several patterns of sample sizes. We let the response probabilities for the first row be 0.6, 0.3, and 0.1. When $r = 3$, we chose the odds ratios ψ_{12}, ψ_{13} to be $(1,1)$, $(2,1)$, or $(4,1)$. When $r = 5$, we chose the odds ratios $\psi_{12}, \psi_{13}, \psi_{14}, \psi_{15}$ to be $(1,1,1,1)$, $(2,1,2,1)$, or $(4,1,4,1)$.

The estimators are considered based on the weights $1/N_k$ or $1/(n_{ik}+n_{hk})$. Let $L(N)$ denote the simpler estimator L_{ih}, which uses only a single $2\times c\times K$ contingency table, with the weight $1/N_k$ and let $L(n)$ denote the simpler estimator L_{ih} with the weight $1/(n_{ik}+n_{hk})$. Similarly, let $\bar{L}(N)$ denote the improved estimator $\bar{L}_{ih}$ with the weight $1/N_k$ and let $\bar{L}(n)$ denote the improved estimator $\bar{L}_{ih}$ with the weight

$1/(n_{ik} + n_{hk})$. For the simpler estimator L_{ih} and the improved estimator $\bar{L}_{ih}$, we study the Pitman ARE for $L(N)$ compared to $L(n)$ and the Pitman ARE for $\bar{L}(N)$ compared to $\bar{L}(n)$, respectively. In addition, we present the Pitman ARE for L_{ih} compared to $\bar{L}_{ih}$ for two different types of weights separately. Also, we show the average Pitman ARE, denoted by 'Avg.', over all possible odds ratios.

Table 4.1 considers the one–stratum case with three nominal row categories. There are two types of sample sizes denoted by I and II. The pattern 'I' means that all a_i are equal for $i \in \{1,2,3\}$. The pattern 'II' means that $a_2 = a_3 = a_1/2$; that is, the marginal counts in the first row are twice the marginal counts in the second row and the marginal counts in the second row are the same as the marginal counts in the third row. In this table, there is not much difference between the estimators, since the average Pitman ARE is close to 1.0. Table 4.2 considers the one–stratum case with five nominal row categories. The sample sizes chosen are similar to those in Table 4.1 by repeating the last two rows in the three nominal row categories case. The table presents the Pitman ARE for ψ_{12}, ψ_{13}, ψ_{14}, ψ_{15}, and ψ_{23} only. The Pitman ARE for the remaining odds ratios (such as ψ_{24}, ψ_{25}, ψ_{34}, ψ_{35}, and ψ_{45}) are similar. The average Pitman ARE is calculated over all those 10 odds ratios. The results in this table are similar to those in Table 4.1.

Table 4.3 considers the three–stratum case with three nominal row categories. The sample size pattern 'I' denotes that $a_{1k} = a_{2k} = a_{3k}$ for all $k \in \{1,2,3\}$, but $N_1 = N_2/2 = N_3$; the pattern 'II' denotes that $N_1 = N_2 = N_3$, but $a_{11}/2 = a_{21} = a_{31}$, $a_{12} = a_{22}/2 = a_{32}$, $a_{13} = a_{23} = a_{33}/2$; and the pattern 'III' denotes the more extreme case in which $10a_{11} = a_{21} = a_{31}$, $a_{12} = a_{22} = 10a_{32}$, $a_{13} = a_{23} = a_{33}/10$. In general, the Pitman ARE does not depend on degree of association; that is, when the pattern of the odds ratios (ψ_{12}, ψ_{13}) is (1,1), (2,1), or (4,1), the asymptotic relative efficiency does not change much. The Pitman ARE can be far away from unity when the

marginal row counts vary drastically across the strata; that is, the results for the sample sizes pattern III are different from those for I and II.

Table 4.4 considers the three–stratum case with five nominal row categories. The results are similar to those in Table 4.3. The study shows that $L(n)$ seems much more efficient than $L(N)$, but $\bar{L}(n)$ seems as efficient as $\bar{L}(N)$. Also, the average asymptotic relative efficiency for $L(n)$ compared to $\bar{L}(n)$ or for $L(N)$ compared to $\bar{L}(N)$ implies that the estimators based on only a part of $r \times c \times K$ contingency tables have at least some loss of efficiency compared to the estimators based on $r \times c \times K$ contingency tables.

We conclude that $\bar{L}_{ih}$ is more efficient than L_{ih}, and $\bar{L}_{ih}$ is much more efficient than L_{ih} when the marginal row counts vary drastically across the strata. For the improved estimator $\bar{L}_{ih}$, two different weights have the similar efficiency. Therefore, it does not hurt to use the improved estimator $\bar{L}$ with the simpler weight $1/N_k$. In summary, the log generalized Mantel–Haenszel–type partial estimators $\{\bar{L}\}$ with the simpler weight are preferable.

The asymptotic relative efficiency discussed is based on the large–stratum situation. However, the Mantel–Haenszel–type estimators are most useful under the sparse–data situation. Now, we discuss the asymptotic relative efficiency for L compared to $\bar{L}$ under the sparse–data case. For $2 \times c \times K$ contingency tables with c nominal response categories, Greenland (1989) showed that $\bar{L}^*$ tends to be much better than L^* when data are sparse. His results also revealed that L^* is less efficient when the marginal column totals vary drastically across the strata. This is similar to our findings, in which L seems to be less efficient when marginal row totals vary drastically across the strata. For $2 \times c \times K$ contingency tables with c nominal response categories and with one observation in each row, the marginal column totals still may vary drastically across the strata. For $r \times c \times K$ contingency tables, the case with one observation in each row is not a situation in which marginal row totals can

vary drastically across the strata. So, we do not expect too much efficiency loss for L in the case of one observation in each row. For the case with extremely unbalanced sample sizes for rows within each stratum, if the row counts vary across the strata, there may be more efficiency loss for L.

The data in Table 4.5, from Madsen (1976), describe an investigation of satisfaction with housing conditions carried out by the Danish Building Research Institute and the Danish Institute of Mental Health Research. We are interested in measuring satisfaction with housing conditions conditional on the degree of contact with the other residents. The satisfaction variable is divided into three groups: low, medium and high. The contact variable is divided into two groups: low and high, and the type of housing is classified into tower blocks, apartment houses, and atrium houses. The high and low contacts are treated as the strata. If we arrange the data into a $r \times c \times K$ contingency table, the row variable is the type of housing with $r = 3$ levels, and the response variable is satisfaction with $c = 3$ levels.

The ML estimators for the model (4.1) can be obtained using Lang's (1992) FORTRAN program. The data in Table 4.5 fit the model very well, since the deviance G^2 is 6.92 with 6 degrees of freedom. Table 4.6 shows the estimates and the corresponding standard errors for the Mantel–Haenszel–type estimators L_{ih} and $\bar{L}_{ih}$ for two different types of weights, and for the ML estimator. Since the data are not sparse, the results are quite similar for the Mantel–Haenszel method and the ML method. Next, we analyze the data using the estimators $\bar{L}_{ih}$ with the weight $1/N_k$. The degree of satisfaction for people living in apartments has no significant difference from people living in atrium houses. The odds of having high satisfaction instead of medium and low satisfaction, or high and medium satisfaction instead of low satisfaction are estimated to be $\exp(0.525) = 1.69$ times higher for people living in tower blocks than in apartments, given each level of contact with other residents. Similarly, the odds of responding high satisfaction instead of medium and low satisfaction, or high and

medium satisfaction instead of low satisfaction are estimated to be $\exp(0.349) = 1.42$ times higher for people living in tower blocks than in atrium houses given each level of contact with other residents.

Table 4.1. The Pitman ARE for L_{ih} compared to $\bar{L}_{ih}$ with different weights, for the one–stratum situation with three nominal row categories and three ordinal response categories.

		ψ_{12}, ψ_{13} 1, 1		ψ_{12}, ψ_{13} 2, 1		ψ_{12}, ψ_{13} 4, 1	
ARE	ψ	I	II	I	II	I	II
$L(N), L(n)$	ψ_{12}	1.0	1.0	1.0	1.0	1.0	1.0
	ψ_{13}	1.0	1.0	1.0	1.0	1.0	1.0
	ψ_{23}	1.0	1.0	1.0	1.0	1.0	1.0
	Avg.	1.0	1.0	1.0	1.0	1.0	1.0
$\bar{L}(N), \bar{L}(n)$	ψ_{12}	1.0	1.0	1.0	1.0	1.0	1.0
	ψ_{13}	1.0	1.0	1.0	1.0	1.0	1.0
	ψ_{23}	1.0	1.0	1.0	1.0	1.0	1.0
	Avg.	1.0	1.0	1.0	1.0	1.0	1.0
$L(n), \bar{L}(n)$	ψ_{12}	1.0	1.0	0.993	0.996	0.981	0.988
	ψ_{13}	1.0	1.0	1.010	1.010	1.027	1.027
	ψ_{23}	1.0	1.0	0.993	0.993	0.981	0.980
	Avg.	1.0	1.0	0.999	1.0	0.996	0.998
$L(N), \bar{L}(N)$	ψ_{12}	1.0	1.0	0.993	0.996	0.981	0.988
	ψ_{13}	1.0	1.0	1.010	1.010	1.027	1.027
	ψ_{23}	1.0	1.0	0.993	0.993	0.981	0.980
	Avg.	1.0	1.0	0.999	1.0	0.996	0.998

Note that I and II denote the sample sizes, where I: a_i are equal for all $i \in \{1, 2, 3\}$ and II: $a_2 = a_3 = a_1/2$.

Table 4.2. The Pitman ARE for L_{ih} compared to $\bar{L}_{ih}$ with different weights, for the one-stratum situation with five nominal row categories and three ordinal response categories.

		$\psi_{12},\psi_{13},\psi_{14},\psi_{15}$		$\psi_{12},\psi_{13},\psi_{14},\psi_{15}$		$\psi_{12},\psi_{13},\psi_{14},\psi_{15}$	
		1,1,1,1		2,1,2,1		4,1,4,1	
ARE	ψ	I	II	I	II	I	II
$L(N),L(n)$	ψ_{12}	1.0	1.0	1.0	1.0	1.0	1.0
	ψ_{13}	1.0	1.0	1.0	1.0	1.0	1.0
	ψ_{14}	1.0	1.0	1.0	1.0	1.0	1.0
	ψ_{15}	1.0	1.0	1.0	1.0	1.0	1.0
	ψ_{23}	1.0	1.0	1.0	1.0	1.0	1.0
	Avg.	1.0	1.0	1.0	1.0	1.0	1.0
$\bar{L}(N),\bar{L}(n)$	ψ_{12}	1.0	1.0	1.0	1.0	1.0	1.0
	ψ_{13}	1.0	1.0	1.0	1.0	1.0	1.0
	ψ_{14}	1.0	1.0	1.0	1.0	1.0	1.0
	ψ_{15}	1.0	1.0	1.0	1.0	1.0	1.0
	ψ_{23}	1.0	1.0	1.0	1.0	1.0	1.0
	Avg.	1.0	1.0	1.0	1.0	1.0	1.0
$L(n),\bar{L}(n)$	ψ_{12}	1.0	1.0	0.993	0.996	0.972	0.978
	ψ_{13}	1.0	1.0	1.013	1.013	1.034	1.034
	ψ_{14}	1.0	1.0	0.993	0.996	0.972	0.978
	ψ_{15}	1.0	1.0	1.013	1.013	1.034	1.034
	ψ_{23}	1.0	1.0	0.993	0.992	0.972	0.972
	Avg.	1.0	1.0	0.999	0.999	0.995	0.996
$L(N),\bar{L}(N)$	ψ_{12}	1.0	1.0	0.993	0.996	0.972	0.978
	ψ_{13}	1.0	1.0	1.013	1.013	1.034	1.034
	ψ_{14}	1.0	1.0	0.993	0.996	0.972	0.978
	ψ_{15}	1.0	1.0	1.013	1.013	1.034	1.034
	ψ_{23}	1.0	1.0	0.993	0.992	0.972	0.972
	Avg.	1.0	1.0	0.999	0.999	0.995	0.996

Note that I and II denote the sample sizes, where I: a_i are equal for all $i \in \{1,\ldots,5\}$ and II: $a_1/2 = a_2 = a_3 = a_4 = a_5$. The table presents the Pitman ARE for ψ_{12}, ψ_{13}, ψ_{14}, ψ_{15}, and ψ_{23} only. The Pitman ARE for the remaining odds ratios (such as ψ_{24}, ψ_{25}, ψ_{34}, ψ_{35}, and ψ_{45}) are similar to those. The average Pitman ARE is calculated over all 10 odds ratios.

Table 4.3. The Pitman ARE for L_{ik} compared to $\bar{L}_{ik}$ with different weights, for the three–stratum situation with three nominal row categories and three ordinal response categories.

		ψ_{12}, ψ_{13} 1, 1			ψ_{12}, ψ_{13} 2, 1			ψ_{12}, ψ_{13} 4, 1		
ARE	ψ	I	II	III	I	II	III	I	II	III
$L(N), L(n)$	ψ_{12}	1.0	0.974	0.686	1.0	0.974	0.688	1.0	0.974	0.689
	ψ_{13}	1.0	0.974	0.971	1.0	0.974	0.971	1.0	0.974	0.971
	ψ_{23}	1.0	0.974	0.988	1.0	0.974	0.992	1.0	0.974	0.994
	Avg.	1.0	0.974	0.882	1.0	0.974	0.884	1.0	0.974	0.884
$\bar{L}(N), \bar{L}(n)$	ψ_{12}	1.0	1.012	1.016	1.0	1.011	1.018	1.0	1.010	1.013
	ψ_{13}	1.0	1.012	1.056	1.0	1.015	1.072	1.0	1.015	1.072
	ψ_{23}	1.0	1.012	0.978	1.0	1.011	0.959	1.0	1.010	0.957
	Avg.	1.0	1.012	1.016	1.0	1.012	1.016	1.0	1.012	1.014
$L(n), \bar{L}(n)$	ψ_{12}	1.0	0.990	1.004	0.993	0.984	1.003	0.981	0.971	0.988
	ψ_{13}	1.0	0.990	0.855	1.010	1.0	0.858	1.027	1.016	0.870
	ψ_{23}	1.0	0.990	0.895	0.993	0.984	0.898	0.981	0.971	0.892
	Avg.	1.0	0.990	0.918	0.999	0.989	0.920	0.996	0.986	0.917
$L(N), \bar{L}(N)$	ψ_{12}	1.0	0.952	0.678	0.993	0.948	0.679	0.981	0.936	0.672
	ψ_{13}	1.0	0.952	0.787	1.010	0.960	0.778	1.027	0.975	0.788
	ψ_{23}	1.0	0.952	0.905	0.993	0.948	0.929	0.981	0.936	0.926
	Avg.	1.0	0.952	0.790	0.999	0.952	0.795	0.996	0.949	0.796

Note that I, II and III denote the sample sizes, where I: $a_{1k} = a_{2k} = a_{3k}$ for all $k \in \{1,2,3\}$ and $N_1 = N_2/2 = N_3$; II: $N_1 = N_2 = N_3$, but $a_{11}/2 = a_{21} = a_{31}$, $a_{12} = a_{22}/2 = a_{32}$, $a_{13} = a_{23} = a_{33}/2$; and III: $10a_{11} = a_{21} = a_{31}$, $a_{12} = a_{22} = 10a_{32}$, $a_{13} = a_{23} = a_{33}/10$.

Table 4.4. The Pitman ARE for L_{ih} compared to $\bar{L}_{ih}$ with different weights, for the three–stratum situation with five nominal row categories and three ordinal response categories.

		$\psi_{12},\psi_{13},\psi_{14},\psi_{15}$ 1,1,1,1			$\psi_{12},\psi_{13},\psi_{14},\psi_{15}$ 2,1,2,1			$\psi_{12},\psi_{13},\psi_{14},\psi_{15}$ 4,1,4,1		
ARE	ψ	I	II	III	I	II	III	I	II	III
$L(N), L(n)$	ψ_{12}	1.0	0.974	0.984	1.0	0.974	0.988	1.0	0.974	0.989
	ψ_{13}	1.0	0.974	0.868	1.0	0.974	0.868	1.0	0.974	0.868
	ψ_{14}	1.0	0.960	0.734	1.0	0.966	0.715	1.0	0.968	0.709
	ψ_{15}	1.0	0.960	0.492	1.0	0.960	0.492	1.0	0.960	0.492
	ψ_{23}	1.0	0.960	0.741	1.0	0.974	0.769	1.0	0.974	0.778
	Avg.	1.0	0.968	0.804	1.0	0.969	0.810	1.0	0.969	0.812
$\bar{L}(N), \bar{L}(n)$	ψ_{12}	1.0	1.020	0.901	1.0	1.019	0.898	1.0	1.019	0.901
	ψ_{13}	1.0	1.020	0.974	1.0	1.022	0.981	1.0	1.022	0.984
	ψ_{14}	1.0	1.011	0.907	1.0	1.013	0.883	1.0	1.014	0.882
	ψ_{15}	1.0	1.011	0.971	1.0	1.012	0.981	1.0	1.012	0.987
	ψ_{23}	1.0	1.020	1.075	1.0	1.019	1.098	1.0	1.019	1.102
	Avg.	1.0	1.013	1.052	1.0	1.013	1.056	1.0	1.013	1.060
$L(n), \bar{L}(n)$	ψ_{12}	1.0	0.976	1.022	0.993	0.969	1.030	0.972	0.949	1.006
	ψ_{13}	1.0	0.976	0.899	1.013	0.988	0.904	1.034	1.008	0.919
	ψ_{14}	1.0	0.996	1.256	0.993	0.991	1.210	0.972	0.972	1.176
	ψ_{15}	1.0	0.996	0.511	1.013	1.008	0.513	1.034	1.029	0.520
	ψ_{23}	1.0	0.976	0.849	0.993	0.969	0.860	0.972	0.949	0.848
	Avg.	1.0	0.990	0.918	0.999	0.989	0.930	0.995	0.985	0.928
$L(N), \bar{L}(N)$	ψ_{12}	1.0	0.932	1.116	0.993	0.926	1.133	0.972	0.907	1.105
	ψ_{13}	1.0	0.932	0.800	1.013	0.941	0.800	1.034	0.960	0.810
	ψ_{14}	1.0	0.945	1.017	0.993	0.945	0.980	0.972	0.928	0.945
	ψ_{15}	1.0	0.945	0.259	1.013	0.956	0.257	1.034	0.976	0.259
	ψ_{23}	1.0	0.932	0.586	0.993	0.926	0.602	0.972	0.907	0.598
	Avg.	1.0	0.947	0.727	0.999	0.946	0.737	0.995	0.942	0.734

Note that I, II, and III denote the sample sizes, where I: $a_{1k} = \cdots = a_{5k}$ for all $k \in \{1,2,3\}$, but $N_1 = N_2/2 = N_3$; II: $N_1 = N_2 = N_3$, $a_{11}/2 = a_{21} = \cdots = a_{51}$, $a_{12} = a_{22}/2 = a_{32} = \cdots = a_{52}$, $a_{13} = a_{23} = a_{33}/2 = a_{43} = a_{53}$; and III: $10a_{11} = a_{21} = a_{31} = 10a_{41} = a_{51}/10$, $a_{12} = a_{22} = 10a_{32} = a_{42}/10 = 10a_{52}$, and $a_{13} = a_{23} = a_{33}/10 = a_{43} = 10a_{53}$. The ways of the presentation of this table is similar to those in Table 4.2.

Table 4.5. The satisfaction with housing conditions with degree of contact with the other residents (Madsen, 1976).

Low contact				High contact			
	Satisfaction				Satisfaction		
Housing	Low	Medium	High	Housing	Low	Medium	High
Tower blocks	65	54	100	Tower blocks	34	47	100
Apartments	130	76	111	Apartments	141	116	191
Atrium houses	27	24	31	Atrium houses	37	55	65

Table 4.6. The Mantel–Haenszel-type estimates and the ML estimates for the data on satisfaction with housing conditions (Table 4.5).

Log of odds ratio	estimate	s. e.	p–value
$L_{12}(n)$	−0.525	0.118	< 0.001
$L_{12}(N)$	−0.522	0.118	< 0.001
$\bar{L}_{12}(n)$	−0.530	0.117	< 0.001
$\bar{L}_{12}(N)$	−0.525	0.117	< 0.001
ML L_{12} $(\hat{\Delta}_1 - \hat{\Delta}_2)$	−0.519	0.117	< 0.001
$L_{13}(n)$	−0.358	0.157	0.022
$L_{13}(N)$	−0.352	0.157	0.025
$\bar{L}_{13}(n)$	−0.353	0.156	0.024
$\bar{L}_{13}(N)$	−0.349	0.156	0.025
ML L_{13} $(\hat{\Delta}_1 - \hat{\Delta}_3)$	−0.360	0.154	0.020
$L_{23}(n)$	0.181	0.138	0.187
$L_{23}(N)$	0.178	0.138	0.196
$\bar{L}_{23}(n)$	0.176	0.138	0.203
$\bar{L}_{23}(N)$	0.176	0.138	0.204
ML L_{23} $(\hat{\Delta}_2 - \hat{\Delta}_3)$	0.159	0.138	0.248

4.3 Ordinal–Ordinal Cumulative Odds Ratios for $r \times c \times K$ Contingency Tables

In this section, we discuss ordinal–ordinal cumulative odds ratios. The 'ordinal–ordinal' case means that the response variable and the explanatory variable are both ordinal. We consider an odds ratio that is 'global' in the columns and is 'local' in the rows. We will explain this odds ratio next. Also, we show that the Mantel–Haenszel–type estimator for this ordinal–ordinal cumulative odds ratio can be obtained by combining the generalized Mantel–Haenszel–type partial estimators for ordinal–nominal cumulative odds ratios provided in Section 4.2.

First, we introduce a proportional odds model to interpret the ordinal–ordinal cumulative odds ratio. Consider a series of K $r \times c$ contingency tables formed by r independent multinomial observations $X_{1jk}, \ldots, X_{rjk}$, $(j = 1, \ldots, c)$ with denominator $n_{1k}, \ldots, n_{rk}$ and probabilities $\pi_{1jk}, \ldots, \pi_{rjk}$, $(j = 1, \ldots, c)$ satisfying $\pi_{i1k} + \cdots + \pi_{ick} = 1$ for $i \in \{1, \ldots, r\}$ and $k \in \{1, \ldots, K\}$. When the explanatory variable Z has r ordinal categories and the response variable Y has c ordinal categories, we can fit a linear trend proportional odds model as follows:

$$\log\left(\frac{\pi^{*}_{ijk}}{1 - \pi^{*}_{ijk}}\right) = \alpha_{jk} - \Delta u_i \,, \tag{4.5}$$

where $i \in \{1, \ldots, r\}$, $u_1 < u_2 < \cdots < u_r$, $j \in \{1, \ldots, c-1\}$, and $k \in \{1, \ldots, K\}$. The $\{u_i\}$ are preassigned scores for the explanatory variable. In the model, each of the $K(c-1)$ cumulative logits is linearly related to the explanatory variable with a common slope Δ.

A special case of the linear trend proportional odds ratio (4.5) assumes equally–spaced preassigned scores. In this section, we consider the equally–spaced case. Without loss of generality, let $u_i = i$ for $i \in \{1, \ldots, r\}$. Then, for any adjacent pair of rows, there is a common cumulative odds ratio $\exp(\Delta)$ across all strata. When

there are only two rows, the linear trend proportional odds model (4.5) is the common proportional odds model (2.1). In the linear trend proportional odds model (4.5), the odds ratio is 'global' in the columns and is 'local' in the rows. The terms 'global' and 'local' mean that for all collapsings of the response into a binary response $\{(\leq j, > j),\ j = 1, \ldots, c-1\}$, the odds ratios are constant for all pairs of adjacent rows; that is,

$$\frac{\pi^*_{ijk}(1-\pi^*_{(i+1)jk})}{(1-\pi^*_{ijk})\pi^*_{(i+1)jk}} = \psi_{i(i+1)} = \exp(\Delta) = \psi_{\mathrm{LC}}\,,$$

for $i \in \{1, \ldots, r\}$, $j \in \{1, \ldots, c-1\}$, and $k \in \{1, \ldots, K\}$. We call ψ_{LC} a *common local cumulative odds ratio.* In general, for a pair of rows i and h, we have an odds ratio

$$\frac{\pi^*_{ijk}(1-\pi^*_{hjk})}{(1-\pi^*_{ijk})\pi^*_{hjk}} = \psi_{ih} = \exp((h-i)\Delta) = (\psi_{\mathrm{LC}})^{h-i}\,.$$

Its interpretation is that the odds of making response $\leq j$ are $\exp((h-i)\Delta)$ times higher at level i of the explanatory variable Z than at level h, for each stratum.

Under the sparse-data situation, the ML estimator for Δ in the linear trend proportional odds model (4.5) may not be consistent. Instead of fitting the model (4.5) by ML, we can summarize the linear trend association between the explanatory variable and the response variable from several $r \times c$ contingency tables using the Mantel–Haenszel method. The Mantel–Haenszel–type estimator is consistent under both sparse–data and large–stratum situations. We propose an estimator for the common local cumulative odds ratio by combining the generalized Mantel–Haenszel–type partial estimators $\{\tilde{L}\}$ discussed in Section 4.2.

Since the estimator $\tilde{L}_{ih}$ is dually consistent for $\log\psi_{ih}$ in the ordinal–nominal cumulative odds ratios case, it is also dually consistent for $(h-i)\Delta$ in the ordinal–ordinal case. Therefore, we propose a dually consistent estimator for Δ using a

weighted average of $\{\bar{L}_{ih}/(h-i)\}$ with weights $\{s_{ih}\}$, denoted as $\bar{\bar{L}}_{\rm LC}$, given by

$$\log(\hat{\psi}_{\rm LC}) = \bar{\bar{L}}_{\rm LC} = \frac{\sum_{i=1}^{r-1}\sum_{h=i+1}^{r} s_{ih}\bar{L}_{ih}/(h-i)}{\sum_{i=1}^{r-1}\sum_{h=i+1}^{r} s_{ih}}, \tag{4.6}$$

where $\bar{L}_{ih} = (L_{i+} - L_{h+})/r$, $L_{ih} = \log(\sum_k \sum_j R_{jk}^{ih}/\sum_k \sum_j S_{jk}^{ih})$, $R_{jk}^{ih} = X_{ijk}^{*}(n_{hk} - X_{hjk}^{*})/N_k$ and $S_{jk}^{ih} = X_{hjk}^{*}(n_{ik} - X_{ijk}^{*})/N_k$. Under the large-stratum limiting model, the proposed MH type estimator (4.6) may have a loss of efficiency, compared to the ML estimator. The weights $\{s_{ih}\}$ may be chosen so that the estimator (4.6) has the smallest loss of efficiency among the class of the MH type estimators. However, the weighted average of the generalized Mantel–Haenszel–type partial estimators, $\hat{\psi}_{\rm LC}$, is dually consistent for the common local cumulative odds ratio, given any chosen weight. It is also asymptotically normally distributed based on the same properties of $\{\bar{L}_{ih}\}$. For simplicity, we suggest using the estimator with equal weights.

As in Section 4.2, we propose dually consistent variance and covariance estimators for $\{L_{ih}\}$ first. Since $\bar{\bar{L}}_{\rm LC}$ is a linear combination of L_{ih}'s, an estimator for $\mathrm{Var}(\bar{\bar{L}}_{\rm LC})$ follows immediately from the variance and covariance for $\{L_{ih}\}$. A dually consistent estimator of $\mathrm{Var}(L_{ih})$ for $i \neq h$ is as in Section 4.2, which is

$$U_{ihh} = \frac{1}{(\hat{\psi}_{\rm LC}^{h-i})^2}\frac{\sum_{k=1}^{K}\hat{\xi}_k^{ih}(\hat{\psi}_{\rm LC}^{h-i})}{[\sum_{k=1}^{K}\sum_{j=1}^{c-1} S_{jk}^{ih}]^2},$$

where $\hat{\xi}_k^{ih}(\psi_{\rm LC}) = \sum_{j=1}^{c-1}\hat{\phi}_{jk}^{ih}(\psi_{\rm LC}) + 2\sum_{j<j'=1}^{c-1}\hat{\phi}_{jj'k}^{ih}(\psi_{\rm LC})$,

$$\begin{aligned}\hat{\phi}_{jk}^{ih}(\psi_{\rm LC}) &= \frac{n_{ik}n_{hk}}{N_k^2}\{\frac{\psi_{\rm LC}^{h-i}(n_{ik}-X_{ijk}^{*})X_{hjk}^{*}}{n_{ik}}[1+(\psi_{\rm LC}^{h-i}-1)\frac{X_{hjk}^{*}}{n_{hk}}] \\ &\quad + \frac{X_{ijk}^{*}(n_{hk}-X_{hjk}^{*})}{n_{hk}}[\psi_{\rm LC}^{h-i}-(\psi_{\rm LC}^{h-i}-1)\frac{X_{ijk}^{*}}{n_{ik}}]\}\end{aligned}$$

$$\begin{aligned}\hat{\phi}_{jj'k}^{ih}(\psi_{\rm LC}) &= \frac{n_{ik}n_{hk}}{N_k^2}\{\frac{\psi_{\rm LC}^{h-i}(n_{ik}-X_{ij'k}^{*})X_{hjk}^{*}}{n_{ik}}[1+(\psi_{\rm LC}^{h-i}-1)\frac{X_{hj'k}^{*}}{n_{hk}}] \\ &\quad + \frac{X_{ijk}^{*}(n_{hk}-X_{hj'k}^{*})}{n_{hk}}[\psi_{\rm LC}^{h-i}-(\psi_{\rm LC}^{h-i}-1)\frac{X_{ij'k}^{*}}{n_{ik}}]\}.\end{aligned}$$

A dually consistent estimator for $\mathrm{Cov}(L_{ih}, L_{ig})$ for $i \neq h \neq g$ is also similar to the estimator in Section 4.2, given by

$$\begin{aligned} U_{ihg} &= \frac{1}{6}[\hat{\mathrm{C}}\mathrm{ov}(L_{ih}, L_{ig}) + \hat{\mathrm{C}}\mathrm{ov}(L_{ig}, L_{ih}) - \hat{\mathrm{C}}\mathrm{ov}(L_{gi}, L_{ih}) - \hat{\mathrm{C}}\mathrm{ov}(L_{hi}, L_{ig}) \\ &\quad + \hat{\mathrm{C}}\mathrm{ov}(L_{gi}, L_{hi}) + \hat{\mathrm{C}}\mathrm{ov}(L_{hi}, L_{gi})]. \end{aligned}$$

Let a and b represent either h or g. Each covariance term in U_{ihg} is as follows:

$$\begin{aligned} \hat{\mathrm{C}}\mathrm{ov}(L_{ia}, L_{ib}) &= \frac{1}{\hat{\psi}_{\mathrm{LC}}^{a-i}\hat{\psi}_{\mathrm{LC}}^{b-i}}\hat{\mathrm{C}}\mathrm{ov}(\hat{\psi}_{ia}, \hat{\psi}_{ib}) \quad \text{, for all } i \neq a \neq b \\ &= \frac{1}{\hat{\psi}_{\mathrm{LC}}^{a-i}\hat{\psi}_{\mathrm{LC}}^{b-i}}[\frac{\sum_{k=1}^{K}\sum_{j=1}^{c-1}\hat{\mathrm{C}}\mathrm{ov}(R_{jk}^{ia} - \psi_{ia}S_{jk}^{ia}, R_{jk}^{ib} - \psi_{ib}S_{jk}^{ib})}{(\sum_k\sum_{j=1}^{c-1}S_{jk}^{ia})(\sum_k\sum_{j=1}^{c-1}S_{jk}^{ib})} \\ &\quad + \frac{\sum_k\sum_{j'\neq j}^{c-1}\hat{\mathrm{C}}\mathrm{ov}(R_{jk}^{ia} - \psi_{ia}S_{jk}^{ia}, R_{j'k}^{ib} - \psi_{ib}S_{j'k}^{ib})}{(\sum_k\sum_{j=1}^{c-1}S_{jk}^{ia})(\sum_k\sum_{j=1}^{c-1}S_{jk}^{ib})}] \\ &= \frac{1}{\hat{\psi}_{\mathrm{LC}}^{a-i}\hat{\psi}_{\mathrm{LC}}^{b-i}}\frac{\sum_k\sum_{j=1}^{c-1}\hat{\phi}_{jk}^{iaib}(\hat{\psi}_{\mathrm{LC}}) + \sum_k\sum_{j\neq j'}^{c-1}\hat{\phi}_{jj'k}^{iaib}(\hat{\psi}_{\mathrm{LC}})}{(\sum_k\sum_{j=1}^{c-1}S_{jk}^{ia})(\sum_k\sum_{j=1}^{c-1}S_{jk}^{ib})}, \end{aligned}$$

$$\hat{\phi}_{jk}^{iaib}(\psi_{\mathrm{LC}}) = \frac{n_{ik}n_{ak}n_{bk}}{N_k^2}[\frac{X_{ijk}^*X_{ajk}^*}{n_{ik}n_{ak}}(\psi_{\mathrm{LC}}^{a-i} - 1) + \frac{X_{ijk}^*}{n_{ik}} - \psi_{\mathrm{LC}}^{a-i}\frac{X_{ajk}^*X_{bjk}^*}{n_{ak}n_{bk}}]$$

$$\begin{aligned} \hat{\phi}_{jj'k}^{iaib}(\psi_{\mathrm{LC}}) &= \frac{n_{ik}\psi_{\mathrm{LC}}^{a-i}}{N_k^2}[X_{ajk}^*(n_{bk} - X_{bj'k}^*)] \text{ if } j < j' \\ &= \frac{n_{ik}\psi_{\mathrm{LC}}^{b-i}}{N_k^2}[X_{bj'k}^*(n_{ak} - X_{ajk}^*)] \text{ if } j > j', \end{aligned}$$

$$\hat{\mathrm{Cov}}(L_{ai}, L_{ib}) = \frac{1}{\hat{\psi}_{\mathrm{LC}}^{i-a}\hat{\psi}_{\mathrm{LC}}^{b-i}}\hat{\mathrm{Cov}}(\hat{\psi}_{ai}, \hat{\psi}_{ib}) \quad \text{, for all } i \neq a \neq b$$

$$= \frac{1}{\hat{\psi}_{\mathrm{LC}}^{i-a}\hat{\psi}_{\mathrm{LC}}^{b-i}}[\frac{\sum_{k=1}^{K}\sum_{j=1}^{c-1}\hat{\mathrm{Cov}}(R_{jk}^{ai} - \psi_{ai}S_{jk}^{ai}, R_{jk}^{ib} - \psi_{ib}S_{jk}^{ib})}{(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ai})(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ib})}$$

$$+ \frac{\sum_k \sum_{j' \neq j}^{c-1} \hat{\mathrm{Cov}}(R_{jk}^{ai} - \psi_{ai}S_{jk}^{ai}, R_{j'k}^{ib} - \psi_{ib}S_{j'k}^{ib})}{(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ai})(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ib})}]$$

$$= \frac{1}{\hat{\psi}_{\mathrm{LC}}^{i-a}\hat{\psi}_{\mathrm{LC}}^{b-i}} \frac{\sum_k \sum_{j=1}^{c-1} \hat{\phi}_{jk}^{aiib}(\hat{\psi}_{\mathrm{LC}}) + \sum_k \sum_{j \neq j'}^{c-1} \hat{\phi}_{jj'k}^{aiib}(\hat{\psi}_{\mathrm{LC}})}{(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ai})(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ib})},$$

$$\hat{\phi}_{jk}^{aiib}(\psi_{\mathrm{LC}}) = \frac{n_{ik}n_{bk}n_{ak}}{N_k^2}[(\psi_{\mathrm{LC}}^{i-a}-1)(\frac{X_{ajk}^* X_{ijk}^*}{n_{ak}n_{ik}} - \frac{X_{ajk}^* X_{ijk}^* X_{bjk}^*}{n_{ik}n_{ak}n_{bk}}) - \psi_{\mathrm{LC}}^{i-a}\frac{X_{ijk}^*}{n_{ik}} + \psi_{\mathrm{LC}}^{i-a}\frac{X_{ijk}^* X_{bjk}^*}{n_{ik}n_{bk}}],$$

$$\hat{\phi}_{jj'k}^{aiib}(\psi_{\mathrm{LC}}) = \frac{n_{ik}}{N_k^2}[-X_{ajk}^*(n_{bk} - X_{bj'k}^*)] \text{ if } j < j'$$

$$= \frac{n_{ik}\psi_{\mathrm{LC}}^{i-a}\psi_{\mathrm{LC}}^{b-i}}{N_k^2}[-(n_{ak} - X_{ajk}^*)X_{bj'k}^*] \text{ if } j > j',$$

$$\hat{\mathrm{Cov}}(L_{ai}, L_{bi}) = \frac{\sum_k \sum_{j=1}^{c-1} \hat{\phi}_{jk}^{aibi}(\hat{\psi}_{\mathrm{LC}}) + \sum_k \sum_{j \neq j'}^{c-1} \hat{\phi}_{jj'k}^{aibi}(\hat{\psi}_{\mathrm{LC}})}{(\hat{\psi}_{ai}\hat{\psi}_{bi})(\sum_k \sum_{j=1}^{c-1} S_{jk}^{bi})(\sum_k \sum_{j=1}^{c-1} S_{jk}^{ai})}, \text{for all } i \neq a \neq b$$

$$\hat{\phi}_{jk}^{aibi}(\psi_{\mathrm{LC}}) = \frac{n_{ik}n_{ak}n_{bk}}{N_k^2}[\psi_{\mathrm{LC}}^{i-a}(\frac{-X_{ajk}^* X_{bjk}^*}{n_{ak}n_{bk}}) + \psi_{\mathrm{LC}}^{i-a}\frac{X_{bjk}^*}{n_{bk}}],$$

$$\hat{\phi}_{jj'k}^{aibi}(\psi_{\mathrm{LC}}) = \frac{n_{ik}\psi_{\mathrm{LC}}^{i-b}}{N_k^2}[X_{ajk}^*(n_{bk} - X_{bj'k}^*)] \text{ if } j < j'$$

$$= \frac{n_{ik}\psi_{\mathrm{LC}}^{i-a}}{N_k^2}[X_{bj'k}^*(n_{ak} - X_{ajk}^*)] \text{ if } j > j'.$$

Similarly, using the same arguments as in Section 4.2, a dually consistent estimator for variance of $\bar{\bar{I}}_{LC}$ is

$$\hat{\mathrm{V}}\mathrm{ar}(\bar{\bar{I}}_{LC}) = \left(\frac{1}{\sum_{i=1}^{r-1}\sum_{h=i+1}^{r} s_{ih}}\right)^2 \times$$

$$\left[\sum_{i=1}^{r-1}\sum_{h=i+1}^{r}\sum_{i'=1}^{r-1}\sum_{h'=i'+1}^{r} s_{ih}s_{i'h'}\hat{\mathrm{C}}\mathrm{ov}(\bar{I}_{ih}, \bar{I}_{i'h'})/[(h-i)(h'-i')]\right],$$

where $\hat{\mathrm{C}}\mathrm{ov}(\bar{I}_{ih}, \bar{I}_{i'h'}) = (U_{ii'}^{+} - U_{ih'}^{+} - U_{hi'}^{+} + U_{hh'}^{+})/r^2$, defined in Section 4.2. If $i = i'$ and $h = h'$, this $\hat{\mathrm{C}}\mathrm{ov}(\bar{I}_{ih}, \bar{I}_{i'h'})$ yields a dually consistent estimator for $\mathrm{Var}(\bar{I}_{ih})$, given by

$$\frac{1}{r^2}(U_{i++} - 2U_{ih}^{+} + U_{h++}).$$

If we reverse the ordering of the response categories, the estimator $\bar{\bar{I}}_{LC}$ equals the negative of the estimator based on the original response. Therefore, the variances of the estimators for the two possible orderings are the same. These arguments are similar to those in Section 4.2. We may use an average of the estimates of the variance from two contingency tables to obtain a symmetric version of the estimate.

4.4 Alternative Ordinal Odds Ratios for $2 \times c \times K$ Contingency Tables

For an ordinal response, instead of fitting a proportional odds model, we can fit an adjacent-category logit model to analyze associations between the response variable and the explanatory variable. Consider a series of K $2 \times c$ contingency tables formed by two independent multinomial observations $(X_{1jk}, X_{2jk}, j = 1, \ldots, c)$, with sample sizes denoted by n_{1k} and n_{2k} and probabilities $(\pi_{1jk}, \pi_{2jk}, j = 1, \ldots, c)$ satisfying $\pi_{i1k} + \cdots + \pi_{ick} = 1$ for $i \in \{1, 2\}$ and $k \in \{1, \ldots, K\}$. We assume that each adjacent pair of columns has the same odds ratio in each $2 \times c$ contingency table.

This assumption corresponds to the adjacent-category logit model

$$\log\left(\frac{\pi_{i,j+1,k}}{\pi_{ijk}}\right) = \alpha_{jk} - \Delta I_{\{i=1\}}\,, i = 1,2,\ j = 1,\ldots,c-1,\ k = 1,\ldots,K. \qquad (4.7)$$

The model implies that the odds ratio $\exp(\Delta)$ has the following structure:

$$\begin{aligned}
\frac{\pi_{11k}\pi_{22k}}{\pi_{21k}\pi_{12k}} &= \frac{\pi_{12k}\pi_{23k}}{\pi_{22k}\pi_{13k}} = \cdots = \frac{\pi_{1,c-1,k}\pi_{2ck}}{\pi_{2,c-1,k}\pi_{1ck}} = \exp(\Delta) = \psi_{\mathrm{L}}\,, \\
\frac{\pi_{11k}\pi_{23k}}{\pi_{21k}\pi_{13k}} &= \frac{\pi_{12k}\pi_{24k}}{\pi_{22k}\pi_{14k}} = \cdots = \frac{\pi_{1,c-2,k}\pi_{2ck}}{\pi_{2,c-2,k}\pi_{1ck}} = \exp(2\Delta) = \psi_{\mathrm{L}}^2\,, \\
&\vdots \\
\frac{\pi_{11k}\pi_{2ck}}{\pi_{21k}\pi_{1ck}} &= \exp((c-1)\Delta) = \psi_{\mathrm{L}}^{c-1}\,,
\end{aligned}$$

for all $k \in \{1,\ldots,K\}$.

In the adjacent-category logit model (4.7), for each stratum, the odds that the response falls at level j instead of $j+1$ for level 1 of the explanatory variable Z is $\exp(\Delta)$ times the odds for level 2 of Z, for all $j \in \{1,\ldots,c-1\}$. For any 2×2 contingency table formed with any adjacent pair of columns, the odds ratio is constant across the $c-1$ adjacent pairs of columns and K strata. For this reason, we call the model (4.7) a *common adjacent-category logit model.* Here, the odds ratio ψ_{L} does not involve the cumulative probabilities, but it consists of single probabilities. The odds ratio is constant across the strata, so we call ψ_{L} a *common local odds ratio.*

When the data are sparse, the unconditional ML estimator for Δ in the common adjacent category logit model (4.7) may not be consistent. This problem is the same as in the logit model and the common proportional odds model (2.1) discussed in Chapter 2. For $2 \times 2 \times K$ contingency tables, a dually consistent common odds ratio estimator can be obtained either by the Mantel–Haenszel method or by fitting the logit model using the conditional ML method. For the common proportional odds model (2.1) in $2 \times c \times K$ contingency tables with an ordinal response, there is no exact

conditional distribution for the common cumulative odds ratio. The exact conditional ML method does not apply in the proportional odds model. For an ordinal response case, Chapter 2 proposed a dually consistent common cumulative odds ratio estimator using the Mantel–Haenszel type method. In this section, we will discuss the Mantel–Haenszel (MH) type estimator for an alternative ordinal odds ratio — the common local odds ratio. For estimating the common local odds ratio, the conditional ML estimator is dually consistent, but it may have an excessive computational burden because of no software available. The MH type estimator for ψ_L is an alternative dually consistent estimator to the conditional ML estimator based on the common adjacent–category logit model (4.7). However, the MH type estimator has a simpler form.

For the common cumulative odds ratio, we proposed the Mantel–Haenszel–type estimate by naively treating all collapsed 2×2 contingency tables as independent samples. Similarly, one way to obtain a Mantel–Haenszel–type estimate of ψ_L is by naively treating each 2×2 contingency table formed by an adjacent pair of columns as an independent stratum. Under the naive assumption, we have $(c-1)K$ strata. Then, a possible estimate is the ordinary Mantel–Haenszel estimate for $K(c-1)$ 2×2 contingency tables. The Mantel–Haenszel estimator is consistent under both limiting cases. The argument follows from Chapter 2. However, there is some loss of efficiency by considering only adjacent pairs of columns, since we lose some information compared to using all possible pairs of columns. Next, we will propose an improved Mantel–Haenszel–type estimator.

The improved Mantel–Haenszel–type estimate follows from the Mickey and Elashoff (1985) generalized Mantel–Haenszel partial estimators, and the dually consistent variance estimator follows from Greenland's (1989) procedure. Since we consider an ordinal response instead of a nominal response, we propose a way to combine their generalized Mantel–Haenszel partial estimators to obtain an improved Mantel–Haenszel–type estimator for the common local odds ratio.

We recall the Mickey and Elashoff estimators from Section 4.2. Let $f_{jj'k} = X_{1jk}X_{2j'k}/N_k$, $F_{jj'} = \sum_k f_{jj'k}$, and $L^*_{jj'} = \log(F_{jj'}/F_{j'j})$. Then, $\exp(L^*_{jj'})$ is the ordinary Mantel–Haenszel estimator for K 2×2 tables, considering columns j and j' alone and using the weight $1/N_k$. From the structure of the common local odds ratio in the common adjacent–category logit model (4.7), $\exp(L^*_{jj'})$ with $j < j' = 1, \ldots, c$, is the Mantel–Haenszel–type estimator for $\exp((j'-j)\Delta)$ or $\psi_{\mathrm{L}}^{(j'-j)}$. Mickey and Elashoff (1985) defined $\psi^*_{jj'}$ to be a partial common odds ratio for the columns j and j', and proposed the generalized Mantel–Haenszel partial estimator for $\psi^*_{jj'}$ as follows:

$$\log(\hat{\psi}^*_{jj'}) = \bar{L}^*_{jj'} = (L^*_{j+} - L^*_{j'+})/c. \tag{4.8}$$

The estimator $\bar{L}^*_{jj'}$ combines information from all possible pairs of columns. Since $L^*_{jj'}$ is dually consistent for $\log\psi^*_{jj'}$, $\bar{L}^*_{jj'}$ is dually consistent as well. The details of the estimator are given in Section 4.2.

Since the estimator $L^*_{jj'}$ is dually consistent for $\log\psi_{\mathrm{L}}^{(j'-j)}$, $\bar{L}^*_{jj'}$ is dually consistent for $\log\psi_{\mathrm{L}}^{(j'-j)}$ as well. We can write $\bar{L}^*_{jj'}/(j'-j) \xrightarrow{p} \Delta = \log\psi_{\mathrm{L}}$. Therefore, we propose a Mantel–Haenszel–type dually consistent estimator using a weighted average of $\{\bar{L}^*_{jj'}/(j'-j)\}$ with weights $\{s^*_{jj'}\}$, denoted by $\bar{\bar{L}}_{\mathrm{L}}$ which is

$$\log(\hat{\psi}_{\mathrm{L}}) = \bar{\bar{L}}_{\mathrm{L}} = \frac{\sum_{j=1}^{c-1}\sum_{j'=j+1}^{c} s^*_{jj'}\bar{L}^*_{jj'}/(j'-j)}{\sum_{j=1}^{c-1}\sum_{j'=j+1}^{c} s^*_{jj'}}. \tag{4.9}$$

The estimator is similar to the estimator for the common local cumulative odds ratio proposed in Section 4.3. Again, the weights $\{s^*_{jj'}\}$ may be chosen so that the estimator

(4.9) has the smallest loss of efficiency among the class of the MH type estimators. For simplicity, we suggest using the estimator with equal weights.

The proposed estimator $\bar{\bar{L}}_{\mathrm{L}}$ (4.9) is asymptotically normal based on the same properties of $\bar{L}^{*}_{ih}$. A dually consistent variance estimator for $\bar{\bar{L}}_{\mathrm{L}}$ can be obtained from a linear combination of dually consistent variance and covariance estimators for $\{L^{*}_{jj'}\}$ by Greenland's (1989) procedure. First, we compute dually consistent variance and covariance estimators for $\{L^{*}_{jj'}\}$ based on the arguments in Chapter 2. A dually consistent estimator for $\mathrm{Var}(L^{*}_{jh})$ follows the estimator of $\mathrm{Var}(\log\hat{\psi})$ (2.10) proposed in Section 2.2, given by

$$\hat{\mathrm{V}}\mathrm{ar}(L^{*}_{jh}) = \frac{1}{(\hat{\psi}_{\mathrm{L}}^{h-j})^2}\frac{\sum_{k=1}^{K}\hat{\varphi}_{k}^{jh}(\hat{\psi}_{\mathrm{L}})}{(F_{hj})^2},$$

where

$$\begin{aligned}\hat{\varphi}_{k}^{jh}(\psi_{\mathrm{L}}) &= \frac{n_{1k}n_{2k}}{N_k^2}\{\frac{\psi_{\mathrm{L}}^{h-j}X_{1hk}X_{2jk}}{n_{1k}}[1+(\psi_{\mathrm{L}}^{h-j}-1)\frac{X_{2jk}}{n_{2k}}]\\ &+ \frac{X_{1jk}X_{2hk}}{n_{2k}}[\psi_{\mathrm{L}}^{h-j}-(\psi_{\mathrm{L}}^{h-j}-1)\frac{X_{1jk}}{n_{1k}}]\}.\end{aligned}$$

Also, since $\mathrm{Var}(L^{*}_{jh})$=$\mathrm{Var}(L^{*}_{hj})$, an invariant estimator for $\mathrm{Var}(L^{*}_{jh})$ is given by

$$U_{jhh} = [\hat{\mathrm{V}}\mathrm{ar}(L^{*}_{jh}) + \hat{\mathrm{V}}\mathrm{ar}(L^{*}_{hj})]/2\ .$$

A dually consistent covariance estimator for L^{*}_{jh} and L^{*}_{jg} for all $j \neq h \neq g$ is given by

$$\begin{aligned}U_{jhg} &= \frac{1}{6}[\hat{\mathrm{C}}\mathrm{ov}(L_{jh},L_{jg}) + \hat{\mathrm{C}}\mathrm{ov}(L_{jg},L_{jh}) - \hat{\mathrm{C}}\mathrm{ov}(L_{gj},L_{jh}) - \hat{\mathrm{C}}\mathrm{ov}(L_{hj},L_{jg})\\ &+\hat{\mathrm{C}}\mathrm{ov}(L_{gj},L_{hj}) + \hat{\mathrm{C}}\mathrm{ov}(L_{hj},L_{gj})].\end{aligned}$$

Let notation a and b represent either g or h. Then, each covariance term in U_{ihg} is as follows:

$$\hat{\mathrm{Cov}}(L_{ja},\, L_{jb}) = \frac{1}{\hat{\psi}_{\mathrm{L}}^{(a-j)}\hat{\psi}_{\mathrm{L}}^{(b-j)}} \frac{\sum_k [X_{1jk}X_{2ak}X_{2bk} + X_{2jk}X_{1ak}X_{1bk}\hat{\psi}_{\mathrm{L}}^{(a-j)}\hat{\psi}_{\mathrm{L}}^{(b-j)}]/N_k^2}{F_{aj}F_{bj}},$$

$$\hat{\mathrm{Cov}}(L_{aj},\, L_{jb}) = \frac{1}{\hat{\psi}_{\mathrm{L}}^{(j-a)}\hat{\psi}_{\mathrm{L}}^{(b-j)}} \frac{\sum_k [-X_{1jk}X_{2ak}X_{2bk}\hat{\psi}_{\mathrm{L}}^{(j-a)} - X_{1ak}X_{1bk}X_{2jk}\hat{\psi}_{\mathrm{L}}^{(b-j)}]/N_k^2}{F_{ja}F_{bj}},$$

$$\hat{\mathrm{Cov}}(L_{aj},\, L_{bj}) = \frac{1}{\hat{\psi}_{\mathrm{L}}^{(j-a)}\hat{\psi}_{\mathrm{L}}^{(j-b)}} \frac{\sum_k [X_{2ak}X_{1jk}X_{2bk}\hat{\psi}_{\mathrm{L}}^{(j-a)}\hat{\psi}_{\mathrm{L}}^{(j-b)} + X_{1ak}X_{1bk}X_{2jk}]/N_k^2}{F_{ja}F_{jb}}.$$

It is easy to show that both U_{jhh} and U_{jhg} are invariant under the interchange of two rows, following the arguments in Section 4.2.

The dually consistent variance estimator for $\bar{\bar{L}}_{\mathrm{L}}$ (4.9) is simply a linear combination of $\{U_{jhg}\}$. We define $U_{jhg} = 0$ if $j = h$ or g, $U_{jh}^{+} = U_{j++}$ if $j = h$, and $U_{jh}^{+} = U_{hj}^{+} = U_{+jh} - U_{jh+} - U_{hj+} + U_{jhh}$ if $j \neq h$. The covariance of $L_{j+}^{\star}$ and $L_{h+}^{\star}$ is

$$\begin{aligned}
&\mathrm{Cov}(L_{j+}^{\star}, L_{h+}^{\star}) = \mathrm{Cov}\Big(\sum_{j'\neq j=1}^{c} L_{jj'}^{\star}, \sum_{h'\neq h=1}^{c} L_{hh'}^{\star}\Big) \\
= \;& \mathrm{Cov}\Big(\sum_{j'\neq j=1}^{c} L_{jj'}^{\star}, L_{hj}^{\star}\Big) + \mathrm{Cov}\Big(L_{jh}^{\star}, \sum_{h'\neq h=1}^{c} L_{hh'}^{\star}\Big) + \sum_{j'\neq j\neq h=1}^{c} \mathrm{Cov}(L_{jj'}^{\star}, L_{hj'}^{\star}) \\
& - \mathrm{Cov}(L_{jh}^{\star}, L_{hj}^{\star}) \\
= \;& -\mathrm{Cov}\Big(\sum_{j'\neq j=1}^{c} L_{jj'}^{\star}, L_{jh}^{\star}\Big) - \mathrm{Cov}\Big(L_{hj}^{\star}, \sum_{h'\neq h=1}^{c} L_{hh'}^{\star}\Big) + \sum_{j'\neq j\neq h=1}^{c} \mathrm{Cov}(L_{j'j}^{\star}, L_{j'h}^{\star}) \\
& + \mathrm{Cov}(L_{jh}^{\star}, L_{jh}^{\star}).
\end{aligned}$$

Then, $\hat{\mathrm{Cov}}(L_{j+}^{\star}, L_{h+}^{\star}) = U_{jh}^{+}$. Consequently, the covariance estimator for $\bar{L}_{jh}^{\star}$ and $\bar{L}_{j'h'}^{\star}$ is

$$\hat{\mathrm{Cov}}(\bar{L}_{jh}^{\star}, \bar{L}_{j'h'}^{\star}) = \frac{1}{c^2}\hat{\mathrm{Cov}}(L_{j+}^{\star} - L_{h+}^{\star}, L_{j'+}^{\star} - L_{h'+}^{\star}) = \frac{1}{c^2}(U_{jj'}^{+} - U_{jh'}^{+} - U_{hj'}^{+} + U_{hh'}^{+}),$$

for $j, j' \in \{1, \ldots, c-1\}$ and $h, h' \in \{1, \ldots, c\}$. If $j = j'$ and $h = h'$, this formula yields a dually consistent estimator for $\mathrm{Var}(\bar{L}^*_{jh})$, which is

$$\hat{\mathrm{V}}\mathrm{ar}(\bar{L}^*_{jh}) = \frac{1}{c^2}(U_{j++} - 2U^+_{jh} + U_{h++}).$$

Therefore, a dually consistent estimator for variance of $\bar{\bar{L}}_{\mathrm{L}}$ is

$$\hat{\mathrm{V}}\mathrm{ar}(\bar{\bar{L}}_{\mathrm{L}}) = \left(\frac{1}{\sum_{j=1}^{c-1}\sum_{j'=j+1}^{c} s^*_{jj'}}\right)^2 \times$$

$$\left[\sum_{j=1}^{c-1}\sum_{h=j+1}^{c}\sum_{j'=1}^{c-1}\sum_{h'=j'+1}^{c} s^*_{jh}s^*_{j'h'}\hat{\mathrm{C}}\mathrm{ov}(\bar{L}^*_{jh}, \bar{L}^*_{j'h'})/[(h-j)(h'-j')]\right].$$

All computations are similar to those in sections 4.2 and 4.3.

In the common adjacent-category logit model (4.7), we assume that the local odds ratios are constant across the strata; that is, there is no interaction between the explanatory variable and the strata. In Chapter 5, we will introduce a more general adjacent-category logit model in which the odds ratios are not constant across the strata.

CHAPTER 5
HETEROGENEOUS ORDINAL ODDS RATIOS ESTIMATION

5.1 Introduction

For a binary response, the ordinary Mantel–Haenszel estimator is widely used when we assume that the associations for K 2×2 contingency tables are the same. This assumption is sometimes reasonable for data such as clinical trial data, in which K medical centers are the K strata and the drugs' effects are approximately the same in each center. For an ordinal response, fitting a proportional odds model is one way to analyze the data. Chapter 2 introduced the common proportional odds model (2.1), assuming that the ordinal associations between the response and the explanatory variable are the same across the K strata.

The proposed Mantel–Haenszel–type estimator for model (2.1) and the estimated variance are obtained based on the assumption that the common cumulative odds ratios are constant across the K strata. In practice, it is sensible to check whether the cumulative odds ratios are constant and to analyze their heterogeneity across the K strata. Section 3.4 discussed several possible tests of the homogeneous cumulative odds ratio assumption, including the Breslow–Day (1980) type test and the goodness–of–fit chi–squared test for the common proportional odds model (2.1). The goodness–of–fit test is adequate only under large sample sizes within each stratum and the Breslow–Day type test is a liberal test under the same limiting situation. However, we are not able to describe the heterogeneity using these methods.

For $2 \times 2 \times K$ contingency tables with a binary response, Liang and Self (1985) proposed a score test for the hypothesis of homogeneous odds ratios, under the assumption that log odds ratios are independent and identically distributed with distribution F. They used the efficient score statistic from the conditional likelihood to establish the test. The conditional distribution is the noncentral hypergeometric distribution, conditioning on the marginal totals for each 2×2 contingency table. The test is appropriate under the sparse–data situation. Also, Liu and Pierce (1993) assumed that the log odds ratios are as a sample from a population with mean θ^* and standard deviation σ^*. Their assumption is similar to the Liang and Self assumption, but their method is more general. Liu and Pierce (1993) were interested in obtaining inferences regarding the parameters (θ^*, σ^*) by direct computation of the conditional likelihood function for (θ^*, σ^*). Using their method, we are able to describe the heterogeneity of the odds ratios in addition to test the assumption.

For $2 \times c \times K$ contingency tables with an ordinal response, we focus on generalizing the Liu and Pierce (1993) procedure. Once we obtain inferences for the heterogeneity of the cumulative odds ratios, we also can test the common cumulative odds ratio assumption. First, we assume that the log cumulative odds ratios are a sample from a population with mean θ and standard deviation σ. The ML estimators for (θ, σ) follow from the likelihood function based on a proportional odds model with a random effect.

Section 5.2 will illustrate the Liu and Pierce (1993) argument for $2 \times 2 \times K$ contingency tables with a binary response. In Section 5.3, we introduce mixed models and provide a possible computation method for $2 \times c \times K$ contingency tables with an ordinal response.

5.2 Heterogeneous Odds Ratios for a Binary Response

A more general model to analyze data in $2\times2\times K$ contingency tables with a binary response is a logit model with interaction between the explanatory variable and the control variable. The model assumes that the association between the response and the explanatory variable may differ from stratum to stratum. We assume that each of the K 2×2 contingency tables is formed by two independent binomial observations (X_{11k}, X_{21k}), with sample sizes denoted by n_{1k} and n_{2k}, for $k = 1, \ldots, K$. The corresponding success probabilities for (X_{11k}, X_{21k}) are (π_{11k}, π_{21k}) for all $k \in \{1, \ldots, K\}$. Then, the model is defined as follows:

$$\log\left(\frac{\pi_{i1k}}{1-\pi_{i1k}}\right) = \beta_k^* + \theta_k^* I_{\{i=1\}} \quad i = 1, 2, \ k = 1, \ldots, K. \tag{5.1}$$

Liu and Pierce (1993) assumed that the log odds ratios $\{\theta_k^*, \ k = 1, 2, \ldots, K\}$ are independent and identically distributed (i. i. d.) with a distribution having mean θ^* and standard deviation σ^*; that is, they treated θ_k^* as a random effect. Because of the random effect θ_k^*, the model (5.1) is a generalized linear mixed model (GLMM). We call the logit model (5.1) a *mixed logit model.* Breslow and Clayton (1993) discussed approximate methods of inference in GLMM's.

In the mixed logit model (5.1), the nuisance parameter β_k^* can be eliminated by conditioning on the marginal totals for each 2×2 contingency table. The cell count X_{11k} follows conditionally the noncentral hypergeometric distribution. The likelihood function of parameters (θ^*, σ^*) for the kth stratum is

$$L_k^*(\theta^*, \sigma^*) \propto \int_{-\infty}^{\infty} L_k^*(t)\phi(t; \theta^*, \sigma^*)dt \ , \tag{5.2}$$

where $L_k^*(\theta_k^*)$ is the likelihood function for the log odds ratio θ_k^* based on the noncentral hypergeometric distribution, and $\phi(\theta_k^*; \theta^*, \sigma^*)$ is the density function of the normal distribution with mean θ^* and standard deviation σ^*. The likelihood function $L_k^*(\theta_k^*)$

has the exponential family form:

$$L_k^*(\theta_k^*) \propto \exp\{X_{11k}\theta_k^* - B_k^*(\theta_k^*)\}\,,$$

where

$$B_k^*(\theta_k^*) = \log \sum_j \binom{n_{1k}}{j} \binom{n_{2k}}{m_{1k}-j} \exp(j\theta_k^*)$$

and the summation is over $\max(0,\, m_{1k} - n_{2k}) \leq j \leq \min(n_{1k},\, m_{1k})$, where $m_{1k} = X_{11k} + X_{21k}$ (Breslow and Day, 1980, p. 125). Thus, the likelihood function over K strata is

$$L(\theta^*, \sigma^*) = \prod_{k=1}^{K} L_k(\theta^*, \sigma^*)\,.$$

Liu and Pierce (1993) pointed out that their approach also applies when $\phi(\theta_k^*; \theta^*, \sigma^*)$ is a density other than normal.

Liu and Pierce (1993) provided an approximation to the integral (5.2), based on Laplace's method (De Bruijn, 1961, p.60; Barndorff–Nielsen and Cox, 1989, p. 169; Tierney and Kadane, 1986). The formula for approximating the likelihood function (5.2) is

$$L_k^*(\theta^*, \sigma^*) \simeq l(\hat{t}_{\theta^*,\sigma^*}; \theta^*, \sigma^*)\sqrt{2\pi/\hat{\gamma}_{\theta^*,\sigma^*}}\,, \tag{5.3}$$

where the first factor is the integrand in (5.2), $l(t; \theta^*, \sigma^*) = L_k^*(t)\phi(t; \theta^*, \sigma^*)$ evaluated at the point $\hat{t}_{\theta^*,\sigma^*}$ that maximizes $l(t; \theta^*, \sigma^*)$ for given (θ^*, σ^*). In the second factor in (5.3),

$$\hat{\gamma}_{\theta^*,\sigma^*} = -\frac{\partial^2}{\partial t^2} \log l(t; \theta^*, \sigma^*)|_{t=\hat{t}_{\theta^*,\sigma^*}} = B_k^{*\prime\prime}(\hat{t}_{\theta^*,\sigma^*}) + \frac{1}{\sigma^{*2}}\,.$$

The Laplace approximation for integrating a function can be used not only for one–dimensional integrals but also for multiple–dimensional integrals. Similarly, Breslow and Clayton (1993) used Laplace approximations for integrating quasi–likelihood functions that are multiple–dimensional integrals in GLMM's.

For evaluating the Laplace approximation, Liu and Pierce (1993) compared it with Gauss–Hermite quadrature. The basic integral for Gauss–Hermite quadrature has the form

$$\int_{-\infty}^{\infty} f(b)\exp(-b^2)db \ ,$$

and it is approximated by

$$\sum_{h=1}^{m} \nu_h f(b_h) \ ,$$

where the nodes b_h are zeros of the mth order Hermite polynomial and the ν_h are corresponding weights. Tables of (b_h, ν_h) for m =2, 3, ..., 10, 12, 16, 20 are given by Abramowitz and Stegun (1972, p. 924). If any integral does not have form $\exp(-b^2)$, a linear transformation can be made so that the integral takes the form $\exp(-b^2)$. In a later article, Liu and Pierce (1994) claimed that order-one Gauss–Hermite quadrature equals the Laplace approximation.

Gauss–Hermite quadrature is often used for numerical integration, because it relates to the Gaussian density. However, it can be used only for one–dimensional integrals. For more general GLMM's that involve several random effects, the Laplace approximation is still appropriate.

5.3 Heterogeneous Odds Ratios for an Ordinal Response

For K $2 \times c$ contingency tables with an ordinal response, we can analyze the data by fitting a proportional odds model. The model (2.1) assumes that the cumulative odds ratios are constant for all $c-1$ collapsed tables and all K strata. However, even if the proportional odds property holds for various collapsings, the cumulative odds ratios may not be the same across the strata. Next, we introduce a more general proportional odds model that has interaction between the explanatory variable and the control variable. This model assumes that the cumulative odds ratios vary across

the strata as follows:

$$\log\left(\frac{\pi^*_{ijk}}{1-\pi^*_{ijk}}\right) = \alpha_j + \beta_k + \theta_k I_{\{i=1\}}\ , i = 1, 2,\ j = 1, \ldots, c-1,\ k = 1, \ldots, K\ , \tag{5.4}$$

where θ_k is the log cumulative odds ratio for table $k = 1, \ldots, K$. Following Liu and Pierce (1993), we assume that $\{\theta_k\}$ are an i. i. d. sample from a normal distribution with mean θ and standard deviation σ. We call model (5.4) a *mixed proportional odds model.*

There are some similarities and differences between the mixed logit model (5.1) and the mixed proportional odds model (5.4). The former model (5.1) is a special case of the latter model (5.4), when the response variable has only two levels. The nuisance parameter β^*_k in the mixed logit model (5.1) can be eliminated by conditioning on the marginal totals for each 2×2 contingency table. The conditional distribution for cell count X_{11k} is the noncentral hypergeometric distribution. However, the nuisance parameters α_j and β_k in the mixed proportional odds model (5.4) cannot be eliminated by conditioning on the marginal totals for each $2 \times c$ contingency table. There is no exact conditional distribution based on the cumulative odds ratios.

From Chapter 2, we assume that each of K $2 \times c$ contingency tables is formed by two independent multinomial observations $(X_{1jk},\ X_{2jk},\ j = 1, \ldots, c)$, with sample sizes denoted by n_{1k} and n_{2k}, $k = 1, \ldots, K$. The corresponding probabilities for $\{X_{ijk},\ i = 1, 2,\ j = 1, \ldots, c,\ k = 1, \ldots, K\}$ are $\{\pi_{ijk},\ i = 1, 2,\ j = 1, \ldots, c,\ k = 1, \ldots, K\}$ with $\pi_{i1k} + \cdots + \pi_{ick} = 1$. Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{c-1})'$, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)'$. The likelihood function based on the mixed proportional odds model (5.4) and the multinomial sampling assumption for the kth stratum is a function of $(\boldsymbol{\alpha},\ \beta_k,\ \theta,\ \sigma)$ with the form

$$L_k(\boldsymbol{\alpha}, \beta_k, \theta, \sigma) \propto \int_{-\infty}^{\infty} \exp\left(\sum_i \sum_j X_{ijk} \log(\pi_{ijk}|t)\right) \phi(t; \theta, \sigma) dt\ , \tag{5.5}$$

where $\phi(\theta_k; \theta, \sigma)$ is the density function of the normal distribution with mean θ and variance σ^2, and $\pi_{ijk}|\theta_k$ is the cell probability based on the mixed proportional odds

model (5.4). The cell probability in (5.5) is

$$\pi_{ijk}|\theta_k = \frac{\exp(\alpha_j + \beta_k + \theta_k I_{\{i=1\}})}{1+\exp(\alpha_j + \beta_k + \theta_k I_{\{i=1\}})} - \frac{\exp(\alpha_{j-1} + \beta_k + \theta_k I_{\{i=1\}})}{1+\exp(\alpha_{j-1} + \beta_k + \theta_k I_{\{i=1\}})},$$

for $i \in \{1,2\}$, $j \in \{1,\ldots,c\}$, $k \in \{1,\ldots,K\}$, where $-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_c = +\infty$. Thus, the likelihood function over the K strata is

$$L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma) = \prod_{k=1}^{K} L_k(\boldsymbol{\alpha},\beta_k,\theta,\sigma). \tag{5.6}$$

Since the likelihood function (5.5) involves only one–dimensional integration, we can use Gauss–Hermite quadrature to perform numerical integration. Also, since we assume that the log cumulative odds ratios θ_k are an i. i. d. sample from the normal distribution, it is natural to approximate the likelihood function (5.5) using Gauss–Hermite quadrature. The likelihood function (5.6) over the K strata is easy to compute after we numerically integrate the likelihood function (5.5). Furthermore, the ML estimators for $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta, \sigma)$ may be obtained by the Newton–Raphson method.

Applying Gauss–Hermite quadrature, the likelihood function (5.5) can be expressed as

$$L_k(\boldsymbol{\alpha},\beta_k,\theta,\sigma) \simeq \sum_{h=1}^{m} \nu_h f_k(b_h;\boldsymbol{\alpha},\beta_k,\theta,\sigma),$$

where

$$\begin{aligned} f_k(b;\boldsymbol{\alpha},\beta_k,\theta,\sigma) &= \exp\left(\sum_i \sum_j X_{ijk} \log\left[\frac{\exp(\alpha_j+\beta_k+[\theta+\sqrt{2}\sigma b]I_{\{i=1\}})}{1+\exp(\alpha_j+\beta_k+[\theta+\sqrt{2}\sigma b]I_{\{i=1\}})}\right.\right. \\ &\quad \left.\left. -\frac{\exp(\alpha_{j-1}+\beta_k+[\theta+\sqrt{2}\sigma b]I_{\{i=1\}})}{1+\exp(\alpha_{j-1}+\beta_k+[\theta+\sqrt{2}\sigma b]I_{\{i=1\}})}\right]\right). \end{aligned}$$

Then the log likelihood function over the K strata (5.6) can be expressed as

$$\log L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma) = \sum_{k=1}^{K} \log L_k(\boldsymbol{\alpha},\beta_k,\theta,\sigma) \simeq \sum_{k=1}^{K} \log\left[\sum_{h=1}^{m} \nu_h f_k(b_h;\boldsymbol{\alpha},\beta_k,\theta,\sigma)\right]. \tag{5.7}$$

Next, we use the Newton–Raphson method to obtain the ML estimators for the log likelihood function (5.7). The method requires an initial guess for the parameters and then generates a sequence of guesses by maximizing second–degree polynomials. Let $\boldsymbol{\lambda}' = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta, \sigma)$. Also, let

$$\boldsymbol{q}' = \left(\frac{\partial \log L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta, \sigma)}{\partial \lambda_1}, \frac{\partial \log L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta, \sigma)}{\partial \lambda_2}, \ldots\right),$$

and let $\boldsymbol{G}$ denote the matrix having entries

$$g_{ab} = \frac{\partial \log L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta, \sigma)}{\partial \lambda_a \partial \lambda_b}.$$

Let $\boldsymbol{q}^{(s)}$ and $\boldsymbol{G}^{(s)}$ be those terms evaluated at $\boldsymbol{\lambda}^{(s)}$, the sth guess for $\hat{\boldsymbol{\lambda}}$. Then, following the Taylor series expansion,

$$\boldsymbol{\lambda}^{(s+1)} = \boldsymbol{\lambda}^{(s)} - (\boldsymbol{G}^{(s)})^{-1}\boldsymbol{q}^{(s)}.$$

We repeat the step until $\boldsymbol{\lambda}^{(s+1)}$ is very close to $\boldsymbol{\lambda}^{(s)}$. Thus, the estimators $\boldsymbol{\lambda}^{(s+1)}$ are the ML estimators for the likelihood function (5.6).

Next, we compute the first and second derivatives for the log likelihood function (5.7) with respect to parameters $\boldsymbol{\lambda}' = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta, \sigma)$. Let the mixed proportional odds model (5.4) have the constraint $\beta_K = 0$. For simplicity, we write $f_k(b_h; \boldsymbol{\alpha}, \beta_k, \theta, \sigma)$ as $f_k(b_h)$. Also, let

$$\begin{aligned}\pi_{ijkh} &= \frac{\exp(\alpha_j + \beta_k + [\theta + \sqrt{2}\sigma b_h] I_{\{i=1\}})}{1 + \exp(\alpha_j + \beta_k + [\theta + \sqrt{2}\sigma b_h] I_{\{i=1\}})} - \frac{\exp(\alpha_{j-1} + \beta_k + [\theta + \sqrt{2}\sigma b_h] I_{\{i=1\}})}{1 + \exp(\alpha_{j-1} + \beta_k + [\theta + \sqrt{2}\sigma b_h] I_{\{i=1\}})} \\ &= \frac{\exp(\vartheta_{ijkh})}{1 + \exp(\vartheta_{ijkh})} - \frac{\exp(\vartheta_{i(j-1)kh})}{1 + \exp(\vartheta_{i(j-1)kh})},\end{aligned}$$

where $\vartheta_{ijkh} = \alpha_j + \beta_k + [\theta + \sqrt{2}\sigma b_h] I_{\{i=1\}}$, and let

$$d(\vartheta_{ijkh}) = \frac{\exp(\vartheta_{ijkh})}{[1 + \exp(\vartheta_{ijkh})]^2},$$

for $i \in \{1,2\}$, $j\{1,\ldots,c\}$, $k \in \{1,\ldots,K\}$, and $h \in \{1,\ldots,m\}$. Then, the first derivatives are as follows:

$$\frac{\partial \log L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma)}{\partial \alpha_j} = \sum_{k=1}^{K}\left(\frac{1}{\sum_{h=1}^{m}\nu_h f_k(b_h)}\right)\left(\sum_{h=1}^{m}\nu_h f_k(b_h)\left[\sum_i\left(\frac{X_{ijk}}{\pi_{ijkh}} - \frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}}\right)d(\vartheta_{ijkh})\right]\right),$$

$$\frac{\partial \log L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma)}{\partial \beta_k} = \left(\frac{1}{\sum_{h=1}^{m}\nu_h f_k(b_h)}\right)\left(\sum_{h=1}^{m}\nu_h f_k(b_h)\left[\sum_i\sum_j\frac{X_{ijk}}{\pi_{ijkh}}\left(d(\vartheta_{ijkh}) - d(\vartheta_{i(j-1)kh})\right)\right]\right),$$

$$\frac{\partial \log L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma)}{\partial \theta} = \sum_k\left(\frac{1}{\sum_{h=1}^{m}\nu_h f_k(b_h)}\right)\left(\sum_{h=1}^{m}\nu_h f_k(b_h)\left[\sum_j\frac{X_{1jk}}{\pi_{1jkh}}\left(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh})\right)\right]\right),$$

$$\frac{\partial \log L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma)}{\partial \sigma} = \sum_k\left(\frac{1}{\sum_{h=1}^{m}\nu_h f_k(b_h)}\right)\left(\sum_{h=1}^{m}\nu_h f_k(b_h)\sqrt{2}b_h\left[\sum_j\frac{X_{1jk}}{\pi_{1jkh}}\left(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh})\right)\right]\right).$$

The second derivatives are as follows:

$$\frac{\partial^2 \log L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma)}{\partial \alpha_j^2}$$

$$= \sum_{k=1}^{K}\left(\frac{-1}{[\sum_{h=1}^{m}\nu_h f_k(b_h)]^2}\right)\left(\sum_{h=1}^{m}\nu_h f_k(b_h)\left[\sum_i(\frac{X_{ijk}}{\pi_{ijkh}}-\frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}})d(\vartheta_{ijkh})\right]\right)^2$$

$$+ \sum_{k=1}^{K}\left(\frac{1}{\sum_{h=1}^{m}\nu_h f_k(b_h)}\right)\left(\sum_{h=1}^{m}\nu_h f_k(b_h)\left[\sum_i(\frac{X_{ijk}}{\pi_{ijkh}}-\frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}})d(\vartheta_{ijkh})\right]^2\right)$$

$$+ \sum_{k=1}^{K}\left(\frac{1}{\sum_{h=1}^{m}\nu_h f_k(b_h)}\right)\left(\sum_{h=1}^{m}\nu_h f_k(b_h)\left[\sum_i(\frac{-X_{ijk}}{\pi_{ijkh}^2}-\frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}^2})d^2(\vartheta_{ijkh})\right.\right.$$

$$\left.\left.+\sum_i(\frac{X_{ijk}}{\pi_{ijkh}}-\frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}})\frac{\exp(\vartheta_{ijkh})-\exp(3\vartheta_{ijkh})}{[1+\exp(\vartheta_{ijkh})]^4}\right]\right),$$

$$\frac{\partial^2 \log L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma)}{\partial \alpha_j \partial \alpha_{j'}}$$

$$= \sum_{k=1}^{K}\left(\frac{-1}{[\sum_{h=1}^{m}\nu_h f_k(b_h)]^2}\right)\left(\sum_{h=1}^{m}\nu_h f_k(b_h)\left[\sum_i(\frac{X_{ij'k}}{\pi_{ij'kh}}-\frac{X_{i(j'+1)k}}{\pi_{i(j'+1)kh}})d(\vartheta_{ij'kh})\right]\right)$$

$$\left(\sum_{h=1}^{m}\nu_h f_k(b_h)\left[\sum_i(\frac{X_{ijk}}{\pi_{ijkh}}-\frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}})d(\vartheta_{ijkh})\right]\right)$$

$$+ \sum_{k=1}^{K}\left(\frac{1}{\sum_{h=1}^{m}\nu_h f_k(b_h)}\right)\left(\sum_{h=1}^{m}\nu_h f_k(b_h)\left[\sum_i(\frac{X_{ij'k}}{\pi_{ij'kh}}-\frac{X_{i(j'+1)k}}{\pi_{i(j'+1)kh}})d(\vartheta_{ij'kh})\right]\right.$$

$$\left.\left[\sum_i(\frac{X_{ijk}}{\pi_{ijkh}}-\frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}})d(\vartheta_{ijkh})\right]\right)$$

$$+ \sum_{k=1}^{K}\left(\frac{1}{\sum_{h=1}^{m}\nu_h f_k(b_h)}\right)\left(\sum_{h=1}^{m}\nu_h f_k(b_h)\left[\sum_i\{(\frac{X_{ijk}}{\pi_{ijkh}^2}d(\vartheta_{i(j-1)kh}))I_{\{j'=j-1\}}\right.\right.$$

$$\left.\left.+(\frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}^2}d(\vartheta_{i(j+1)kh}))I_{\{j'=j+1\}}\right\}d(\vartheta_{ijkh})\right]\right),$$

$$
\begin{aligned}
&\frac{\partial^2 \log L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta, \sigma)}{\partial \alpha_j \partial \beta_k} \\
= \; & \left(\frac{-1}{[\sum_{h=1}^{m} \nu_h f_k(b_h)]^2}\right)\left(\sum_{h=1}^{m} \nu_h f_k(b_h)\left[\sum_i \sum_{l=1}^{c} \frac{X_{ilk}}{\pi_{ilkh}}(d(\vartheta_{ilkh}) - d(\vartheta_{i(l-1)kh}))\right]\right) \\
& \left(\sum_{h=1}^{m} \nu_h f_k(b_h)\left[\sum_i (\frac{X_{ijk}}{\pi_{ijkh}} - \frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}}) d(\vartheta_{ijkh})\right]\right) \\
+ \; & \left(\frac{1}{\sum_{h=1}^{m} \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^{m} \nu_h f_k(b_h)\left[\sum_i \sum_{l=1}^{c} \frac{X_{ilk}}{\pi_{ilkh}}(d(\vartheta_{ilkh}) - d(\vartheta_{i(l-1)kh}))\right]\right. \\
& \left.\left[\sum_i (\frac{X_{ijk}}{\pi_{ijkh}} - \frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}}) d(\vartheta_{ijkh})\right]\right) \\
+ \; & \left(\frac{1}{\sum_{h=1}^{m} \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^{m} \nu_h f_k(b_h)\left[\sum_i \{\frac{-X_{ijk}}{\pi_{ijkh}^2}(d(\vartheta_{ijkh}) - d(\vartheta_{i(j-1)kh}))\right.\right. \\
& + \frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}^2}(d(\vartheta_{i(j+1)kh}) - d(\vartheta_{ijkh}))\} d(\vartheta_{ijkh}) \\
& \left.\left. + \sum_i (\frac{X_{ijk}}{\pi_{ijkh}} - \frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}})(\frac{\exp(\vartheta_{ijkh}) - \exp(3\vartheta_{ijkh})}{[1 + \exp(\vartheta_{ijkh})]^4})\right]\right),
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial^2 \log L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma)}{\partial\alpha_j\partial\theta} \\
&= \sum_k \left(\frac{-1}{[\sum_{h=1}^m \nu_h f_k(b_h)]^2}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\left[\sum_{l=1}^e \frac{X_{1lk}}{\pi_{1lkh}}(d(\vartheta_{1lkh}) - d(\vartheta_{1(l-1)kh}))\right]\right) \\
&\quad \left(\sum_{h=1}^m \nu_h f_k(b_h)\left[\sum_i (\frac{X_{ijk}}{\pi_{ijkh}} - \frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}})d(\vartheta_{ijkh})\right]\right) \\
&+ \sum_k \left(\frac{1}{\sum_{h=1}^m \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\left[\sum_{l=1}^e \frac{X_{1lk}}{\pi_{1lkh}}(d(\vartheta_{1lkh}) - d(\vartheta_{1(l-1)kh}))\right]\right. \\
&\quad \left.\left[\sum_i (\frac{X_{ijk}}{\pi_{ijkh}} - \frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}})d(\vartheta_{ijkh})\right]\right) \\
&+ \sum_k \left(\frac{1}{\sum_{h=1}^m \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\left[\{\frac{-X_{1jk}}{\pi_{1jkh}^2}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))\right.\right. \\
&\quad + \frac{X_{1(j+1)k}}{\pi_{1(j+1)kh}^2}(d(\vartheta_{1(j+1)kh}) - d(\vartheta_{1jkh}))\}d(\vartheta_{1jkh}) \\
&\quad \left.\left. + (\frac{X_{1jk}}{\pi_{1jkh}} - \frac{X_{1(j+1)k}}{\pi_{1(j+1)kh}})(\frac{\exp(\vartheta_{1jkh}) - \exp(3\vartheta_{1jkh})}{[1+\exp(\vartheta_{1jkh})]^4})\right]\right),
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial^2 \log L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma)}{\partial\alpha_j\partial\sigma} \\
&= \sum_k \left(\frac{-1}{[\sum_{h=1}^m \nu_h f_k(b_h)]^2}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\sqrt{2}b_h\left[\sum_{l=1}^c \frac{X_{1lk}}{\pi_{1lkh}}(d(\vartheta_{1lkh}) - d(\vartheta_{1(l-1)kh}))\right]\right) \\
&\quad \left(\sum_{h=1}^m \nu_h f_k(b_h)\left[\sum_i (\frac{X_{ijk}}{\pi_{ijkh}} - \frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}})d(\vartheta_{ijkh})\right]\right) \\
&+ \sum_k \left(\frac{1}{\sum_{h=1}^m \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\sqrt{2}b_h\left[\sum_{l=1}^c \frac{X_{1lk}}{\pi_{1lkh}}(d(\vartheta_{1lkh}) - d(\vartheta_{1(l-1)kh}))\right]\right. \\
&\quad \left.\left[\sum_i (\frac{X_{ijk}}{\pi_{ijkh}} - \frac{X_{i(j+1)k}}{\pi_{i(j+1)kh}})d(\vartheta_{ijkh})\right]\right) \\
&+ \sum_k \left(\frac{1}{\sum_{h=1}^m \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\sqrt{2}b_h\left[\{\frac{-X_{1jk}}{\pi_{1jkh}^2}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))\right.\right. \\
&\quad \left. + \frac{X_{1(j+1)k}}{\pi_{1(j+1)kh}^2}(d(\vartheta_{1(j+1)kh}) - d(\vartheta_{1jkh}))\}d(\vartheta_{1jkh})\right. \\
&\quad \left.\left. + (\frac{X_{1jk}}{\pi_{1jkh}} - \frac{X_{1(j+1)k}}{\pi_{1(j+1)kh}})(\frac{\exp(\vartheta_{1jkh}) - \exp(3\vartheta_{1jkh})}{[1+\exp(\vartheta_{1jkh})]^4})\right]\right), \\
&\frac{\partial^2 \log L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma)}{\partial\beta_k^2} \\
&= \left(\frac{-1}{[\sum_{h=1}^m \nu_h f_k(b_h)]^2}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\left[\sum_i\sum_j \frac{X_{ijk}}{\pi_{ijkh}}(d(\vartheta_{ijkh}) - d(\vartheta_{i(j-1)kh}))\right]\right)^2 \\
&+ \left(\frac{1}{\sum_{h=1}^m \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\left[\sum_i\sum_j \frac{X_{ijk}}{\pi_{ijkh}}(d(\vartheta_{ijkh}) - d(\vartheta_{i(j-1)kh}))\right]^2\right) \\
&+ \left(\frac{1}{\sum_{h=1}^m \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\left[\sum_i\sum_j \frac{-X_{ijk}}{\pi_{ijkh}^2}(d(\vartheta_{ijkh}) - d(\vartheta_{i(j-1)kh}))^2\right.\right. \\
&\quad \left.\left. + \sum_i\sum_j \frac{X_{ijk}}{\pi_{ijkh}}(\frac{\exp(\vartheta_{ijkh}) - \exp(3\vartheta_{ijkh})}{[1+\exp(\vartheta_{ijkh})]^4} - \frac{\exp(\vartheta_{i(j-1)kh}) - \exp(3\vartheta_{i(j-1)kh})}{[1+\exp(\vartheta_{i(j-1)kh})]^4})\right]\right), \\
&\frac{\partial^2 \log L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma)}{\partial\beta_k\partial\beta_{k'}} = 0\,,
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial^2 \log L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta, \sigma)}{\partial \beta_k \partial \theta} \\
= &\left(\frac{-1}{[\sum_{h=1}^{m} \nu_h f_k(b_h)]^2}\right)\left(\sum_{h=1}^{m} \nu_h f_k(b_h)\left[\sum_j \frac{X_{1jk}}{\pi_{1jkh}}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))\right]\right) \\
&\left(\sum_{h=1}^{m} \nu_h f_k(b_h)\left[\sum_i \sum_j \frac{X_{ijk}}{\pi_{ijkh}}(d(\vartheta_{ijkh}) - d(\vartheta_{i(j-1)kh}))\right]\right) \\
+ &\left(\frac{1}{\sum_{h=1}^{m} \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^{m} \nu_h f_k(b_h)\left[\sum_j \frac{X_{1jk}}{\pi_{1jkh}}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))\right]\right. \\
&\left.\left[\sum_i \sum_j \frac{X_{ijk}}{\pi_{ijkh}}(d(\vartheta_{ijkh}) - d(\vartheta_{i(j-1)kh}))\right]\right) \\
+ &\left(\frac{1}{\sum_{h=1}^{m} \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^{m} \nu_h f_k(b_h)\left[\sum_j \frac{-X_{1jk}}{\pi_{1jkh}^2}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))^2\right.\right. \\
&\left.\left. + \sum_j \frac{X_{1jk}}{\pi_{1jkh}}\left(\frac{\exp(\vartheta_{1jkh}) - \exp(3\vartheta_{1jkh})}{[1+\exp(\vartheta_{1jkh})]^4} - \frac{\exp(\vartheta_{1(j-1)kh}) - \exp(3\vartheta_{1(j-1)kh})}{[1+\exp(\vartheta_{1(j-1)kh})]^4}\right)\right]\right),
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial^2 \log L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma)}{\partial\beta_k\partial\sigma} \\
&= \left(\frac{-1}{[\sum_{h=1}^m \nu_h f_k(b_h)]^2}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\sqrt{2}b_h\left[\sum_j \frac{X_{1jk}}{\pi_{1jkh}}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))\right]\right) \\
&\quad \left(\sum_{h=1}^m \nu_h f_k(b_h)\left[\sum_i\sum_j \frac{X_{ijk}}{\pi_{ijkh}}(d(\vartheta_{ijkh}) - d(\vartheta_{i(j-1)kh}))\right]\right) \\
&+ \left(\frac{1}{\sum_{h=1}^m \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\sqrt{2}b_h\left[\sum_j \frac{X_{1jk}}{\pi_{1jkh}}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))\right]\right. \\
&\quad \left.\left[\sum_i\sum_j \frac{X_{ijk}}{\pi_{ijkh}}(d(\vartheta_{ijkh}) - d(\vartheta_{i(j-1)kh}))\right]\right) \\
&+ \left(\frac{1}{\sum_{h=1}^m \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\sqrt{2}b_h\left[\sum_j \frac{-X_{1jk}}{\pi_{1jkh}^2}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))^2\right.\right. \\
&\quad \left.\left.+\sum_j \frac{X_{1jk}}{\pi_{1jkh}}(\frac{\exp(\vartheta_{1jkh}) - \exp(3\vartheta_{1jkh})}{[1+\exp(\vartheta_{1jkh})]^4} - \frac{\exp(\vartheta_{1(j-1)kh}) - \exp(3\vartheta_{1(j-1)kh})}{[1+\exp(\vartheta_{1(j-1)kh})]^4})\right]\right),
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial^2 \log L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma)}{\partial\theta^2} \\
&= \sum_k\left(\frac{-1}{[\sum_{h=1}^m \nu_h f_k(b_h)]^2}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\left[\sum_j \frac{X_{1jk}}{\pi_{1jkh}}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))\right]\right)^2 \\
&+ \sum_k\left(\frac{1}{\sum_{h=1}^m \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\left[\sum_j \frac{X_{1jk}}{\pi_{1jkh}}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))\right]^2\right) \\
&+ \sum_k\left(\frac{1}{\sum_{h=1}^m \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\left[\sum_j \frac{-X_{1jk}}{\pi_{1jkh}^2}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))^2\right.\right. \\
&\quad \left.\left.+\sum_j \frac{X_{1jk}}{\pi_{1jkh}}(\frac{\exp(\vartheta_{1jkh}) - \exp(3\vartheta_{1jkh})}{[1+\exp(\vartheta_{1jkh})]^4} - \frac{\exp(\vartheta_{1(j-1)kh}) - \exp(3\vartheta_{1(j-1)kh})}{[1+\exp(\vartheta_{1(j-1)kh})]^4})\right]\right),
\end{aligned}
$$

$$\frac{\partial^2 \log L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma)}{\partial\theta\partial\sigma}$$

$$= \sum_k \left(\frac{-1}{[\sum_{h=1}^m \nu_h f_k(b_h)]^2}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\sqrt{2}b_h\left[\sum_j \frac{X_{1jk}}{\pi_{1jkh}}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))\right]\right)$$

$$\left(\sum_{h=1}^m \nu_h f_k(b_h)\left[\sum_j \frac{X_{1jk}}{\pi_{1jkh}}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))\right]\right)$$

$$+ \sum_k \left(\frac{1}{\sum_{h=1}^m \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\sqrt{2}b_h\left[\sum_j \frac{X_{1jk}}{\pi_{1jkh}}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))\right]^2\right)$$

$$+ \sum_k \left(\frac{1}{\sum_{h=1}^m \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\sqrt{2}b_h\left[\sum_j \frac{-X_{1jk}}{\pi_{1jkh}^2}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))^2\right.\right.$$

$$\left.\left. + \sum_j \frac{X_{1jk}}{\pi_{1jkh}}\left(\frac{\exp(\vartheta_{1jkh}) - \exp(3\vartheta_{1jkh})}{[1+\exp(\vartheta_{1jkh})]^4} - \frac{\exp(\vartheta_{1(j-1)kh}) - \exp(3\vartheta_{1(j-1)kh})}{[1+\exp(\vartheta_{1(j-1)kh})]^4}\right)\right]\right),$$

$$\frac{\partial^2 \log L(\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\sigma)}{\partial\sigma^2}$$

$$= \sum_k \left(\frac{-1}{[\sum_{h=1}^m \nu_h f_k(b_h)]^2}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)\sqrt{2}b_h\left[\sum_j \frac{X_{1jk}}{\pi_{1jkh}}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))\right]\right)^2$$

$$+ \sum_k \left(\frac{1}{\sum_{h=1}^m \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)2b_h^2\left[\sum_j \frac{X_{1jk}}{\pi_{1jkh}}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))\right]^2\right)$$

$$+ \sum_k \left(\frac{1}{\sum_{h=1}^m \nu_h f_k(b_h)}\right)\left(\sum_{h=1}^m \nu_h f_k(b_h)2b_h^2\left[\sum_j \frac{-X_{1jk}}{\pi_{1jkh}^2}(d(\vartheta_{1jkh}) - d(\vartheta_{1(j-1)kh}))^2\right.\right.$$

$$\left.\left. + \sum_j \frac{X_{1jk}}{\pi_{1jkh}}\left(\frac{\exp(\vartheta_{1jkh}) - \exp(3\vartheta_{1jkh})}{[1+\exp(\vartheta_{1jkh})]^4} - \frac{\exp(\vartheta_{1(j-1)kh}) - \exp(3\vartheta_{1(j-1)kh})}{[1+\exp(\vartheta_{1(j-1)kh})]^4}\right)\right]\right).$$

Based on those first and second derivatives, the ML estimators should be obtainable.

There is one problem with the ML estimators for the mixed proportional odds model (5.4) under the sparse-data situation. Since we cannot eliminate the nuisance parameter β_k, the number of parameters $\{\beta_k\}$ grows with the sample size under the sparse-data situation. Thus, the ML estimators for θ and σ may not be consistent.

Even though we may not be able to use this model when the data are sparse, the mixed model is still useful in practice when the sample sizes are large within each stratum.

For an ordinal response, we can alternatively analyze the data by fitting an adjacent–category logit model, discussed in Section 4.4. Section 4.4 introduced the common adjacent–category logit model (4.7), assuming that the local odds ratios are the same for each adjacent pair of columns for each $2 \times c$ contingency table. However, a more general model assumes that the local odds ratios may vary across the strata. We consider a more general adjacent–category logit model with interaction between the explanatory variable and the control variable as follows:

$$\log\left(\frac{\pi_{i,j+1,k}}{\pi_{ijk}}\right) = \alpha_j + \beta_k - \theta_k I_{\{i=1\}}, i = 1, 2, j = 1, \ldots, c-1, \ k = 1, \ldots, K. \tag{5.8}$$

If we assume that the log local odds ratio θ_k are i. i. d. normal distributed with mean θ and variance σ^2, the model (5.8) is a type of GLMM, called a *mixed adjacent–category logit model.*

For the mixed adjacent–category logit model (5.8), we can eliminate the nuisance parameters α_j and β_k by conditioning on the marginal totals for each $2 \times c$ contingency table. The conditional distribution is the noncentral multiple hypergeometric distribution (Plackett, 1981, p. 48). Using the similar procedure as for the mixed proportional odds model, the conditional ML estimators can be obtained. Unlike the ML estimators for the mixed proportional odds model, the conditional ML estimators for the mixed adjacent–category logit model are appropriate under both limiting situations.

Conditioning on the row and column marginal totals, Plackett (1981) provided the noncentral multiple hypergeometric distribution for $r \times c$ contingency tables with nominal rows and columns. The odds ratios discussed by Plackett can be written as functions of the local odds ratio θ_k in each stratum, when the explanatory variable

has two levels. Therefore, conditioning on the marginal totals in the k stratum, the cell counts have the noncentral multiple hypergeometric distribution involving a parameter θ_k. This conditional likelihood function has an exponential family form. Since θ_k is normally distributed with mean θ and σ, the marginal likelihood function of parameters (θ, σ) for the kth stratum can be obtained by integrating over θ_k. Also, the likelihood function over K strata is a product of the K likelihood functions for all strata. We then numerically integrate the likelihood function using Gauss–Hermite quadrature. Furthermore, the dually consistent conditional ML estimators for (θ, σ) can be conducted by the Newton–Raphson method.

This chapter does not provide any example, since computing work has not been finished. In summary, for both mixed models, we can perform the inferences regarding the mean and variance for the K ordinal odds ratios based on numerical integration. For the mixed proportional odds model (5.4), there is no conditional ML estimators for the parameters. The ML estimators may have good asymptotics only under large sample sizes within each stratum. However, for the mixed adjacent-category logit model (5.8), the conditional ML estimators can be obtained. They may be appropriate under both limiting models.

CHAPTER 6
CONCLUSION

6.1 Summary of Results

For an ordinal response case, the ML estimator of an ordinal odds ratio for a proportional odds model may not be a good estimator when the number of strata increases with the sample size. If the response is binary, the conditional ML estimator and the ordinary Mantel–Haenszel estimator for the common odds ratio are dually consistent estimators. For the common cumulative odds ratio, there is no conditional ML estimator based on the common proportional odds model, since the exact conditional distribution does not exist. The McCullagh and Nelder (1989) pseudo 'conditional likelihood' estimator cannot be used for most cases in the sparse–data situation. Therefore, we proposed the Mantel–Haenszel type estimator to estimate the common cumulative odds ratio. The proposed estimator is consistent when the sample size within each stratum becomes large and the number of strata is fixed and when the number of strata increases with the sample size. We also proposed the dually consistent variance estimator for our Mantel–Haenszel type estimator. In addition, we showed that the generalization of the Mantel–Haenszel test of conditional independence under an ordinal response case is the Mantel (1963) score test.

Chapter 3 evaluated the ordinal odds ratio estimator. A simulation study showed that the proposed estimator tends to be much better than the ML estimator under the sparse–data situation. When each stratum consists of a matched pair of subjects, the expected value of the ML estimator for the log common cumulative odds ratio is about twice the true parameter, but the expected value of the proposed estimator is very

close to the true parameter. On the other hand, under the large–stratum situation, the ML estimator is asymptotically efficient. So, the proposed MH type estimator may have some efficiency loss under Model I. However, based on comparisons of the Pitman asymptotic relative efficiency, the proposed estimator does not usually have much loss of efficiency compared to the ML estimator.

The ordinary Mantel–Haenszel method assumes that the odds ratio is constant across the strata. There are several tests for testing the homogeneous odds ratio assumption. Similarly, the proposed Mantel–Haenszel–type estimator assumes that the cumulative odds ratio is constant across the strata. To obtain the proposed estimator, we collapsed the responses into a binary response ($\leq j, > j$) and treated the odds ratios as constant across the collapsings and the strata. For testing the assumption of a common cumulative odds ratio, under the large–stratum situation, we can test goodness–of–fit by fitting a proportional odds model with no interaction between the explanatory variable and the control variable. If the model fits well, it is plausible that the odds ratios are constant across the collapsings and the strata. Under the sparse–data situation, we proposed a test of the common cumulative odds ratio across the $c-1$ collapsings. This test is not only appropriate under the sparse–data limiting model, but also under the large–stratum limiting model.

For a binary response case, another test of the assumption is the Breslow–Day test, which is valid only under the large–stratum situation. However, some authors (Tarone, 1985; Yanagawa and Fujii, 1990) pointed out that the Breslow–Day test statistic is not asymptotically chi–squared distributed, and they corrected it so the new test statistic is asymptotically chi–squared. In Section 3.4, we generalized the Breslow–Day test statistic to the ordinal response case.

Chapter 4 extended our estimator to either r nominal rows or ordinal rows. When the explanatory variable has r nominal categories, we can fit a proportional odds model with row effects; when the explanatory variable has r ordinal categories, we

can fit a linear trend proportional odds model. Based on these models, we estimated the partial common cumulative odds ratios and the common local cumulative odds ratio using the generalized Mantel–Haenszel–type method. In addition, we proposed the dually consistent variance and covariance estimators for them.

For $2 \times c \times K$ contingency tables with an ordinal response, we can analyze the data by fitting an adjacent–category logit model instead of a proportional odds model. By conditioning on the margins for each $2 \times c$ contingency table, the counts follow the noncentral multiple hypergeometric distribution based on the common local odds ratio. Similar to the binary response case, the conditional ML estimator for the common local odds ratio is dually consistent, but it may have an excessive computational burden because of no software available. Section 4.4 proposed an alternative dually consistent estimator, which has a simple form, by the Mantel–Haenszel type method. The section also presented a dually consistent variance estimator for it.

Instead of testing homogeneous cumulative odds ratios, we could describe their heterogeneity. Chapter 5 proposed a proportional odds model with a random effect, which is one type of generalized linear mixed model (GLMM). Our model assumed that the log cumulative odds ratios for table $k = 1, \ldots, K$ are an i. i. d. sample from a normal distribution with unknown mean and variance. The ML estimators for the mean and variance can be obtained by maximizing the likelihood function for the mixed proportional odds model. The likelihood function involves one–dimensional integration. We used the Gauss–Hermite quadrature to perform numerical integration. For the mixed proportional odds model, since the nuisance parameters cannot be eliminated, the ML estimators are only good for the large–strata situation.

An alternative GLMM model for an ordinal response is a mixed adjacent–category logit model. Chapter 5 also proposed a mixed adjacent–category logit model assuming that the log local odds ratios for table $k = 1, \ldots, K$ are an i. i. d. sample from a normal distribution with unknown mean and variance. In this model, we can eliminate

the nuisance parameters by conditioning on the margins for each $2 \times c$ contingency table. Then, the conditional ML estimators for the mean and variance will be good under both limiting situations.

6.2 Future Research

Some topics in the dissertation can be extended. First at all, the two types of generalized linear mixed model introduced in Chapter 5 need more word. In practice, it is important to analyze the heterogeneity of log ordinal odds ratios and to measure their mean and variance. For the mixed proportional odds model (5.4), we are currently working on the computational part. Consequently, we shall construct a test of the homogeneity based on it. Similarly, for the mixed adjacent–category logit model (5.8), we need further discussions of the conditional ML estimators for the mean and variance of log local odds ratios. Since we cannot eliminate the nuisance parameters for the mixed proportional odds model, the likelihood function involves the mean, variance, and the nuisance parameters. However, the conditional likelihood function for the mixed adjacent–category logit model is only a function of the mean and variance. It should be easier to obtain the ML estimators based on it than for the mixed proportional odds model. Also, the conditional ML estimators are appropriate under both limiting situations.

For the binary response case, the Breslow–Day test statistic is not asymptotically chi–squared distributed. Tarone (1985) corrected it so the new test statistic is asymptotically chi–squared. A future research topic is to generalize the corrections proposed by Yanagawa and Fujii (1990) to obtain an asymptotic chi–squared Breslow–Day type test.

Section 4.4 discussed the common adjacent–category logit model (4.7) and proposed a dually consistent Mantel–Haenszel type estimator for the common local odds ratio. Conditioning on the row and column marginal totals, the cell counts have the

noncentral multiple hypergeometric distribution for the common local odds ratio. In addition to the Mantel–Haenszel type estimator, the conditional ML estimator is also dually consistent. Unlike the common local odds ratio, the common cumulative odds ratio discussed in Chapter 2 has no exact conditional ML estimator. The ML estimator for it is consistent only under the large–stratum case. Therefore, Chapter 3 described the loss of efficiency for the MH type estimator compared to the ML estimator only under the large–stratum limiting model. In the future, we can evaluate the loss of efficiency for the Mantel–Haenszel type estimator compared to the conditional ML estimator under both limiting models based on the common local odds ratio.

For $r \times c \times K$ contingency tables with r ordinal rows and c ordinal columns, Section 4.3 introduced a linear trend proportional odds model. It assumes that an odds ratio is 'global' in the columns and is 'local' in the rows. When the preassigned scores are equally–spaced in the model (4.5), the cumulative odds ratio is constant across the strata and for all adjacent pairs of rows. We then proposed a dually consistent Mantel–Haenszel type estimator for the common local cumulative odds ratio in the section. Alternatively, we can consider an odds ratio that is 'global' in the columns and in the rows. When we collapse the response into a binary response ($\leq j$, $> j$) and collapse the r levels of the explanatory variable into two levels ($\leq i$, $> i$), the collapsed odds ratio is constant for any i and j within each stratum. This type of odds ratio is referred to as a *global* odds ratio (Dale, 1986, and Agresti, 1984, p. 20). The common global odds ratio assumes that the global odds ratio is constant across the strata. For the common global odds ratio, it is reasonable to assume that each of the K $r \times c$ contingency tables is formed by one multinomial observation. This is different from the assumption given in Section 4.3 that each of the K $r \times c$ tables is formed by r independent multinomial observations. The Mantel–Haenszel type method can be generalized to the common global odds ratio case. However, in

the future, the dually consistent variance estimator for it needs more work without assuming r independent rows.

In Section 4.3, we proposed the weighted average of the generalized Mantel–Haenszel–type partial estimators (4.6) for the common local cumulative odds ratio in $r \times c \times K$ contingency tables with ordinal rows and columns. Under the large–stratum limiting model, the proposed MH type estimator may have a loss of efficiency, compared to the ML estimator. We claimed that the weights can be chosen so that the estimator (4.6) has the smallest loss of efficiency among the class of the MH type estimators. An additional future problem is to find the optimal weights. Similarly, we will work on the optimal weights for the MH type common local odds ratio estimator (4.9) proposed in Section 4.4.

APPENDIX A
CALCULATIONS FOR SECTION 2.4

In this appendix, we will prove the equality, $\psi^2 t_k^2/\omega_k = \lim_{N_+\to\infty}[(\sum_j D_{jk} + 2\sum_{j<j'} D_{jj'k})/N_+]$ discussed in Section 2.4. This equality is used to establish the proofs of the consistency for our proposed variance estimator under the large-stratum situation. First, we recall some notations.

Let $t_k = (a_{1k}^{-1} + a_{2k}^{-1})^{-1}[\sum_{j=1}^{c-1}(1-\pi_{1jk}^*)\pi_{2jk}^*]$, where $N_+ a_{ik} = n_{ik}$ for $i \in \{1,2\}$. Let $\omega_k^{-1} = (1/a_{1k})[\sum_{a=1}^{c}\pi_{1ak}\eta_{1ak}^2] + (1/a_{2k})[\sum_{b=1}^{c}\pi_{2bk}\eta_{2bk}^2]$, where $\eta_{1ak} = (1/A_k)[\sum_{b=a+1}^{c}(b-a)\pi_{2bk}] - (1/B_k)[\sum_{b=1}^{a-1}(a-b)\pi_{2bk}]$, and $\eta_{2bk} = (1/A_k)[\sum_{a=1}^{b-1}(b-a)\pi_{1ak}] - (1/B_k)[\sum_{a=b+1}^{c}(a-b)\pi_{1ak}]$, $A_k = \sum_{a<b}^{c}(b-a)\pi_{1ak}\pi_{2bk}$, $B_k = \sum_{a>b}^{c}(a-b)\pi_{1ak}\pi_{2bk}$, $a \in \{1,\ldots,c\}$, $b \in \{1,\ldots,c\}$, and $k \in \{1,\ldots,K\}$. Also,

$$D_{jk} = \frac{n_{1k}n_{2k}}{N_k^2}\{n_{2k}[1+(\psi-1)\pi_{2jk}^*]\psi\pi_{2jk}^*(1-\pi_{1jk}^*) + n_{1k}[\psi-(\psi-1)\pi_{1jk}^*]\pi_{1jk}^*(1-\pi_{2jk}^*)]\},$$

$$\begin{aligned} D_{jj'k} &= \frac{n_{1k}n_{2k}}{N_k^2}\{n_{2k}[1+(\psi-1)(1-\pi_{2j'k}^*)]\psi\pi_{1j'k}^*\pi_{2jk}^* \\ &\quad + n_{1k}[\psi-(\psi-1)(1-\pi_{1j'k}^*)]\pi_{1jk}^*\pi_{2j'k}^*\}, \end{aligned}$$

where $\pi_{ijk}^* = \pi_{i(j+1)k} + \cdots + \pi_{ick} = 1 - \pi_{ijk}^*$. We will prove that

$$\frac{\psi^2 t_k^2}{\omega_k} = \lim_{N_+\to\infty}\frac{\sum_j D_{jk} + 2\sum_{j<j'} D_{jj'k}}{N_+}.$$

Proof. Let $n_{1k}/N_k \to l_k$, $n_{2k}/N_k \to 1-l_k$, and $N_k/N_+ \to r_k$. The term a_{1k} can be written as $l_k r_k$ and the term a_{2k} can be written as $(1-l_k)r_k$. For convenience, we

drop the subscript k, referring to a single $2 \times c$ table. First, we simplify the terms $\lim_{N_+ \to \infty}(D_j/N_+)$ and $\lim_{N_+ \to \infty}(D_{jj'}/N_+)$. Since

$$\psi = \frac{\pi^*_{1j}(1-\pi^*_{2j})}{\pi^*_{2j}(1-\pi^*_{1j})}, \quad \text{and} \quad \psi - 1 = \frac{\pi^*_{1j}-\pi^*_{2j}}{\pi^*_{2j}(1-\pi^*_{1j})}, \tag{A.1}$$

we have

$$\begin{aligned}
&\lim_{N_+ \to \infty} \frac{D_j}{N_+} \\
&= r(1-l)^2 l[1+(\psi-1)\pi^*_{2j}]\psi\pi^*_{2j}(1-\pi^*_{1j}) + r(1-l)l^2[\psi-(\psi-1)\pi^*_{1j}]\pi^*_{1j}(1-\pi^*_{2j}) \\
&= r(1-l)^2 l[1+(\psi-1)\pi^*_{2j}]\pi^*_{1j}(1-\pi^*_{2j}) + r(1-l)l^2[\psi\pi^*_{1j}(1-\pi^*_{2j}) \\
&\quad -(\pi^*_{1j})^2(1-\pi^*_{2j})\psi + \psi\pi^*_{1j}\pi^*_{2j}(1-\pi^*_{1j})] \\
&= r(1-l)^2 l[\pi^*_{1j}(1-\pi^*_{2j}) - \pi^*_{2j}\pi^*_{1j}(1-\pi^*_{2j}) + \psi\pi^*_{2j}\pi^*_{1j}(1-\pi^*_{2j})] \\
&\quad +r(1-l)l^2\psi[\pi^*_{1j}(1-\pi^*_{2j})(1-\pi^*_{1j}) + \pi^*_{1j}\pi^*_{2j}(1-\pi^*_{1j})] \\
&= r(1-l)^2 l[\pi^*_{1j}(1-\pi^*_{2j})^2 + \psi\pi^*_{1j}\pi^*_{2j}(1-\pi^*_{2j})] \\
&\quad +r(1-l)l^2\psi[\pi^*_{1j}(1-\pi^*_{1j})(1-\pi^*_{2j}+\pi^*_{2j})] \\
&= r(1-l)^2 l[\psi\pi^*_{2j}(1-\pi^*_{2j})(1-\pi^*_{1j}) + \psi\pi^*_{1j}\pi^*_{2j}(1-\pi^*_{2j})] + r(1-l)l^2\psi[\pi^*_{1j}(1-\pi^*_{1j})] \\
&= r(1-l)^2 l\psi\pi^*_{2j}(1-\pi^*_{2j}) + r(1-l)l^2\psi\pi^*_{1j}(1-\pi^*_{1j}),
\end{aligned}$$

and

$$
\begin{aligned}
&\lim_{N_+\to\infty}\frac{D_{jj'}}{N_+}\\
&= r(1-l)^2 l[1+(\psi-1)(1-\underline{\pi}^*_{2j'})]\psi\underline{\pi}^*_{1j'}\pi^*_{2j}\\
&\quad +r(1-l)l^2[\psi-(\psi-1)(1-\underline{\pi}^*_{1j'})]\pi^*_{1j}\underline{\pi}^*_{2j'}\\
&= r(1-l)^2 l[\psi\underline{\pi}^*_{1j'}\pi^*_{2j}+(\psi-1)\psi\pi^*_{2j'}\underline{\pi}^*_{1j'}\pi^*_{2j}]\\
&\quad +r(1-l)l^2[\pi^*_{1j}\underline{\pi}^*_{2j'}\psi-(\psi-1)\pi^*_{1j'}\pi^*_{1j}\underline{\pi}^*_{2j'}]\\
&= r(1-l)^2 l[\psi(1-\pi^*_{1j'})\pi^*_{2j}+\psi(\pi^*_{1j'}-\pi^*_{2j'})\pi^*_{2j}]\\
&\quad +r(1-l)l^2[\psi\pi^*_{1j}(1-\pi^*_{2j'})-(\psi-1)\psi\pi^*_{2j'}\underline{\pi}^*_{1j'}\pi^*_{1j}]\\
&= r(1-l)^2 l[\psi\pi^*_{2j}(1-\pi^*_{2j'})]+r(1-l)l^2[\psi\pi^*_{1j}(1-\pi^*_{2j'})-\psi\pi^*_{1j}(\pi^*_{1j'}-\pi^*_{2j'})]\\
&= r(1-l)^2 l\psi\pi^*_{2j}(1-\pi^*_{2j'})+r(1-l)l^2\psi\pi^*_{1j}(1-\pi^*_{1j'}).
\end{aligned}
$$

The term $\psi^2 t^2/\omega$ can be partitioned into two parts. We write

$$
\begin{aligned}
\frac{\psi^2 t^2}{\omega} &= \psi^2\{(a_1^{-1}+a_2^{-1})^{-1}[\sum_{j=1}^{c-1}\pi^*_{2j}(1-\pi^*_{1j})]\}^2\omega^{-1}\\
&= \psi^2\{rl(1-l)[\sum_{j=1}^{c-1}\pi^*_{2j}(1-\pi^*_{1j})]\}^2\omega^{-1}=\sum_{j=1}^{c-1}V_j+V_{\mathrm{R}},
\end{aligned}
$$

where $\sum_{j=1}^{c-1}V_j$ denotes the first part, and $V_{\mathrm{R}}=\psi^2t^2/\omega-\sum_{j=1}^{c-1}V_j$ denotes the remaining part. Also, the first part V_j is defined as

$$
V_j=\psi^2[rl(1-l)\pi^*_{2j}(1-\pi^*_{1j})]^2\omega_j^{-1},
$$

where

$$
\omega_j^{-1}=\frac{1}{rl\pi^*_{1j}(1-\pi^*_{1j})}+\frac{1}{r(1-l)\pi^*_{2j}(1-\pi^*_{2j})}.
$$

Using the properties of the common odds ratios (A.1), the term V_j equals

$$\begin{aligned} V_j &= \psi^2 rl(1-l)^2 \frac{1}{\pi^*_{1j}}(\pi^*_{2j})^2(1-\pi^*_{1j}) + \psi^2 rl^2(1-l)\frac{1}{1-\pi^*_{2j}}\pi^*_{2j}(1-\pi^*_{1j})^2 \\ &= \psi rl(1-l)^2\pi^*_{2j}(1-\pi^*_{2j}) + \psi rl^2(1-l)\pi^*_{1j}(1-\pi^*_{1j}). \end{aligned}$$

From the previous calculation of $\lim_{N_+\to\infty}(D_j/N_+)$, we have $V_j = \lim_{N_+\to\infty}(D_j/N_+)$. Then, we conclude that

$$\sum_{j=1}^{c-1} V_j = \sum_{j=1}^{c-1} \lim_{N_+\to\infty} \frac{D_j}{N_+}.$$

Therefore, in order to prove $\psi^2 t_k^2/\omega_k = \lim_{N_+\to\infty}[(\sum_j D_{jk} + 2\sum_{j<j'} D_{jj'k})/N_+]$, we have to prove the remaining part $V_R = \lim_{N_+\to\infty} 2\sum_{j<j'}(D_{jj'}/N_+)$; that is, we have to prove that

$$V_R = 2rl(1-l)^2\psi \sum_{j=1}^{c-2} \pi^*_{2j} \sum_{j'=j+1}^{c-1} (1-\pi^*_{2j'}) + 2rl^2(1-l)\psi \sum_{j=1}^{c-2} \pi^*_{1j} \sum_{j'=j+1}^{c-1} (1-\pi^*_{1j'}).$$

The term $V_{\rm R}$ equals

$$
\begin{aligned}
V_{\rm R} &= \frac{\psi^2 l^2}{\omega} - \sum_{j=1}^{c-1} V_j \\
&= \psi^2 r^2 l^2 (1-l)^2 \omega^{-1} [\sum_{j=1}^{c-1} \pi_{2j}^* (1-\pi_{1j}^*)]^2 - \psi^2 r^2 l^2 (1-l)^2 [\sum_{j=1}^{c-1} (\pi_{2j}^*)^2 (1-\pi_{1j}^*)^2 \omega_j^{-1}] \\
&= \psi^2 r^2 l^2 (1-l)^2 \{ \left[\frac{1}{a_1} \sum_{a=1}^{c} \pi_{1a} \left(\frac{\sum_{b=a+1}^{c} (b-a)\pi_{2b}}{\sum_{a<b} (b-a)\pi_{1a}\pi_{2b}} - \frac{\sum_{b=1}^{a-1} (a-b)\pi_{2b}}{\sum_{a>b} (a-b)\pi_{1a}\pi_{2b}} \right)^2 \right] \\
&\quad + \left[\frac{1}{a_2} \sum_{b=1}^{c} \pi_{2b} \left(\frac{\sum_{a=1}^{b-1} (b-a)\pi_{1a}}{\sum_{a<b} (b-a)\pi_{1a}\pi_{2b}} - \frac{\sum_{a=b+1}^{c} (a-b)\pi_{1a}}{\sum_{a>b} (a-b)\pi_{1a}\pi_{2b}} \right)^2 \right] \} \\
&\quad \times \{ \sum_{j=1}^{c-1} \pi_{2j}^* (1-\pi_{1j}^*) \}^2 \\
&\quad - \psi r l (1-l)^2 \sum_{j=1}^{c-1} \pi_{2j}^* (1-\pi_{2j}^*) - \psi r l^2 (1-l) \sum_{j=1}^{c-1} \pi_{1j}^* (1-\pi_{1j}^*) \\
&= \psi^2 r^2 l^2 (1-l)^2 \{ \left[\frac{1}{rl} \sum_{a=1}^{c} \pi_{1a} \left(\frac{\sum_{b=a+1}^{c} (b-a)\pi_{2b}}{\sum_{a<b} (b-a)\pi_{1a}\pi_{2b}} - \frac{\sum_{b=1}^{a-1} (a-b)\pi_{2b}}{\sum_{a>b} (a-b)\pi_{1a}\pi_{2b}} \right)^2 \right] \\
&\quad + \left[\frac{1}{r(1-l)} \sum_{b=1}^{c} \pi_{2b} \left(\frac{\sum_{a=1}^{b-1} (b-a)\pi_{1a}}{\sum_{a<b} (b-a)\pi_{1a}\pi_{2b}} - \frac{\sum_{a=b+1}^{c} (a-b)\pi_{1a}}{\sum_{a>b} (a-b)\pi_{1a}\pi_{2b}} \right)^2 \right] \} \\
&\quad \times \{ \sum_{a>b} (a-b)\pi_{1a}\pi_{2b} \}^2 \\
&\quad - \psi r l (1-l)^2 \sum_{j=1}^{c-1} \pi_{2j}^* (1-\pi_{2j}^*) - \psi r l^2 (1-l) \sum_{j=1}^{c-1} \pi_{1j}^* (1-\pi_{1j}^*) \\
&= \psi^2 r l (1-l)^2 K_1 + \psi^2 r l^2 (1-l) K_2 \,.
\end{aligned}
$$

Next, we will prove that $K_1 = (2/\psi) \sum_{j=1}^{c-2} \pi_{2j}^* \sum_{j'=j+1}^{c-1} (1-\pi_{2j'}^*)$.

First, we compute some equalities of equations that are useful for the proof of $K_1 = (2/\psi)\sum_{j=1}^{c-2}\pi_{2j}^{*}\sum_{j'=j+1}^{c-1}(1-\pi_{2j'}^{*})$. The equations are as follows:

$$\pi_{1a}(\sum_{b=a+1}^{c}(b-a)\pi_{2b})^2 = \pi_{1a}(\pi_{2(a+1)}+2\pi_{2(a+2)}+\cdots+(c-a)\pi_{2c})^2$$

$$= \pi_{1a}[(1-\pi_{1a}^{*})+\cdots+(1-\pi_{2(c-1)}^{*})]^2$$

$$= \pi_{1a}\sum_{b=a}^{c-1}(1-\pi_{2b}^{*})^2+2\pi_{1a}\sum_{b=a}^{c-2}\sum_{b'=b+1}^{c-1}(1-\pi_{2b}^{*})(1-\pi_{2b'}^{*}),$$

$$\pi_{1a}[\sum_{b=a+1}^{c}(b-a)\pi_{2b}][\sum_{b=1}^{a-1}(a-b)\pi_{2b}]$$

$$= \pi_{1a}[\pi_{2(a+1)}+2\pi_{2(a+2)}+\cdots+(c-a)\pi_{2c}][(a-1)\pi_{21}+(a-2)\pi_{22}+\cdots+\pi_{2(a-1)}]$$

$$= \pi_{1a}[\sum_{b=a}^{c-1}(1-\pi_{2b}^{*})][\sum_{b=1}^{a-1}\pi_{2b}^{*}],$$

$$\pi_{1a}[\sum_{b=1}^{a-1}(a-b)\pi_{2b}]^2 = \pi_{1a}[(a-1)\pi_{21}+(a-2)\pi_{22}+\cdots+\pi_{2(a-1)}]^2$$

$$= \pi_{1a}\sum_{b=1}^{a-1}(\pi_{2b}^{*})^2+2\pi_{1a}\sum_{b=1}^{a-2}\sum_{b'=b+1}^{a-1}\pi_{2b}^{*}\pi_{2b'}^{*},$$

and $A = \psi B$. Using the above equations, K_1 can be written as

$$
\begin{aligned}
K_1 &= \frac{B^2}{A^2}\Bigg[\pi_{11}\sum_{j=1}^{c-1}(1-\pi_{2j}^*)^2 + 2\pi_{11}\sum_{j=1}^{c-2}\sum_{j'=j+1}^{c-1}(1-\pi_{2j}^*)(1-\pi_{2j'}^*) \\
&+ \pi_{12}\sum_{j=2}^{c-1}(1-\pi_{2j}^*)^2 + 2\pi_{12}\sum_{j=2}^{c-2}\sum_{j'=j+1}^{c-1}(1-\pi_{2j}^*)(1-\pi_{2j'}^*) \\
&- 2\psi\pi_{12}\sum_{j=2}^{c-1}(1-\pi_{2j}^*)\pi_{21} + \psi^2\pi_{12}\pi_{21}^2 + \pi_{13}\sum_{j=3}^{c-1}(1-\pi_{2j}^*)^2 \\
&+ 2\pi_{13}\sum_{j=3}^{c-2}\sum_{j'=j+1}^{c-1}(1-\pi_{2j}^*)(1-\pi_{2j'}^*) - 2\psi\pi_{13}(\sum_{j=3}^{c-1}(1-\pi_{2j}^*))(\sum_{j=1}^{3-1}\pi_{2j}^*) \\
&+ \psi^2\pi_{13}\sum_{j=1}^{3-1}(\pi_{2j}^*)^2 + 2\psi^2\pi_{13}\sum_{j=1}^{3-2}\sum_{j'=j+1}^{3-1}\pi_{2j}^*\pi_{2j'}^* + \cdots + \psi^2\pi_{1c}\sum_{j=1}^{c-1}(\pi_{2j}^*)^2 \\
&+ 2\psi^2\pi_{1c}\sum_{j=1}^{c-2}\sum_{j'=j+1}^{c-1}\pi_{2j}^*\pi_{2j'}^*\Bigg] - \frac{1}{\psi}\sum_{j=1}^{c-1}\pi_{2j}^*(1-\pi_{2j}^*) \\
&= \frac{1}{\psi^2}\Bigg[\sum_{j=1}^{c-1}\pi_{1j}\sum_{j'=j}^{c-1}(1-\pi_{2j'}^*)^2 + 2\sum_{h=1}^{c-2}\pi_{1h}\sum_{j=h}^{c-2}\sum_{j'=j+1}^{c-1}(1-\pi_{2j}^*)(1-\pi_{2j'}^*) \\
&- 2\psi\sum_{j=2}^{c-1}\pi_{1j}(\sum_{j'=j}^{c-1}(1-\pi_{2j'}^*))(\sum_{j'=1}^{j-1}\pi_{2j'}^*) + \psi^2\sum_{j=2}^{c}\pi_{1j}\sum_{j'=1}^{j-1}(\pi_{2j'}^*)^2 \\
&+ 2\psi^2\sum_{h=3}^{c}\pi_{1h}\sum_{j=1}^{h-2}\sum_{j'=j+1}^{h-1}\pi_{2j}^*\pi_{2j'}^*\Bigg] - \frac{1}{\psi}\sum_{j=1}^{c-1}\pi_{2j}^*(1-\pi_{2j}^*).
\end{aligned}
$$

Again, there are some equalities of equations as follows:

$$\sum_{j=1}^{c-1}\pi_{1j}\sum_{j'=j}^{c-1}(1-\pi^*_{2j'})^2$$

$$= \pi_{11}\sum_{j=1}^{c-1}(1-\pi^*_{2j})^2+\pi_{12}\sum_{j=2}^{c-1}(1-\pi^*_{2j})^2+\cdots+\pi_{1(c-1)}\sum_{j=c-1}^{c-1}(1-\pi^*_{2j})^2$$

$$= \pi^*_{11}(1-\pi^*_{21})^2+\pi^*_{12}(1-\pi^*_{22})^2+\pi^*_{13}(1-\pi^*_{23})^2+\cdots+\pi^*_{1(c-1)}(1-\pi^*_{2(c-1)})^2$$

$$= \sum_{j=1}^{c-1}\pi^*_{1j}(1-\pi^*_{2j})^2,$$

$$\sum_{h=1}^{c-2}\pi_{1h}\sum_{j=h}^{c-2}\sum_{j'=j+1}^{c-1}(1-\pi^*_{2j})(1-\pi^*_{2j'})$$

$$= \pi_{11}\sum_{j=1}^{c-2}\sum_{j'=j+1}^{c-1}(1-\pi^*_{2j})(1-\pi^*_{2j'})+\pi_{12}\sum_{j=2}^{c-2}\sum_{j'=j+1}^{c-1}(1-\pi^*_{2j})(1-\pi^*_{2j'})+\ldots$$

$$+\pi_{1(c-2)}\sum_{j=c-2}^{c-2}\sum_{j'=j+1}^{c-1}(1-\pi^*_{2j})(1-\pi^*_{2j'})$$

$$= \pi^*_{11}(1-\pi^*_{21})\sum_{j'=2}^{c-1}(1-\pi^*_{2j'})+\pi^*_{12}(1-\pi^*_{22})\sum_{j'=3}^{c-1}(1-\pi^*_{2j'})+\ldots$$

$$+\pi^*_{1(c-2)}(1-\pi^*_{2(c-2)})(1-\pi^*_{2(c-1)})$$

$$= \sum_{j=1}^{c-2}\pi^*_{1j}(1-\pi^*_{2j})\sum_{j'=j+1}^{c-1}(1-\pi^*_{2j'}),$$

$$\sum_{j=2}^{c-1}\pi_{1j}\left(\sum_{j'=j}^{c-1}(1-\pi^*_{2j'})\right)\left(\sum_{j'=1}^{j-1}\pi^*_{2j'}\right)$$

$$= \pi_{12}\left(\sum_{j=2}^{c-1}(1-\pi^*_{2j})\right)\pi^*_{21}+\pi_{13}\left(\sum_{j=3}^{c-1}(1-\pi^*_{2j})\right)\left(\sum_{j'=1}^{2}\pi^*_{2j'}\right)+\dots$$

$$+\pi_{1(c-1)}(1-\pi^*_{2(c-1)})(\sum_{j'=1}^{c-2}\pi^*_{2j'})$$

$$= \left[\pi^*_{12}(1-\pi^*_{22})(\pi^*_{21})+\pi^*_{13}(1-\pi^*_{23})(\pi^*_{21}+\pi^*_{22})+\dots+\pi^*_{1(c-1)}(1-\pi^*_{2(c-1)})(\sum_{j'=1}^{c-2}\pi^*_{2j'})\right]$$

$$-\left[\pi^*_{11}\pi^*_{21}\sum_{j=2}^{c-1}(1-\pi^*_{2j})+\pi^*_{12}\pi^*_{22}\sum_{j=3}^{c-1}(1-\pi^*_{2j})+\dots+\pi^*_{1(c-2)}\pi^*_{2(c-2)}\sum_{j=c-1}^{c-1}(1-\pi^*_{2j})\right]$$

$$= \sum_{j=2}^{c-1}\pi^*_{1j}(1-\pi^*_{2j})(\sum_{j'=1}^{j-1}\pi^*_{2j'})-\sum_{j=1}^{c-2}\pi^*_{1j}\pi^*_{2j}\sum_{j'=j+1}^{c-1}(1-\pi^*_{2j'}),$$

$$\sum_{j=2}^{c}\pi_{1j}\sum_{j'=1}^{j-1}(\pi^*_{2j'})^2$$

$$= \pi_{12}(\pi^*_{21})^2+\pi_{13}[(\pi^*_{21})^2+(\pi^*_{22})^2]+\dots+\pi_{1c}[(\pi^*_{21})^2+\dots+(\pi^*_{2(c-1)})^2]$$

$$= (\pi^*_{21})^2(\pi_{12}+\pi_{13}+\dots+\pi_{1c})+(\pi^*_{22})^2(\pi_{13}+\dots+\pi_{1c})+\dots+(\pi^*_{2(c-1)})^2\pi_{1c}$$

$$= \sum_{j=1}^{c-1}(\pi^*_{2j})^2(1-\pi^*_{1j}),$$

$$\sum_{h=3}^{c}\pi_{1h}\sum_{j=1}^{h-2}\sum_{j'=j+1}^{h-1}\pi^*_{2j}\pi^*_{2j'}$$

$$= \pi_{13}\pi^*_{21}\pi^*_{22}+\pi_{14}(\pi^*_{21}\pi^*_{22}+\pi^*_{21}\pi^*_{23}+\pi^*_{22}\pi^*_{23})+\dots$$

$$= (1-\pi^*_{12})\pi^*_{22}\pi^*_{21}+(1-\pi^*_{13})\pi^*_{23}(\pi^*_{21}+\pi^*_{22})+\dots$$

$$+(1-\pi^*_{1(c-1)})\pi^*_{2(c-1)}(\pi^*_{21}+\dots+\pi^*_{2(c-2)})$$

$$= \sum_{j=2}^{c-1}(1-\pi^*_{1j})\pi^*_{2j}\sum_{j'=1}^{j-1}\pi^*_{2j'}.$$

Then, we use the above equalities and the properties of the common cumulative odds ratios (A.1) to simplify K_1 as

$$
\begin{aligned}
K_1 &= \frac{1}{\psi^2}\Bigg[\sum_{j=1}^{c-1}\pi^*_{1j}(1-\pi^*_{2j})^2 + 2\sum_{j=1}^{c-2}\pi^*_{1j}(1-\pi^*_{2j})\sum_{j'=j+1}^{c-1}(1-\pi^*_{2j'}) \\
&- 2\psi\sum_{j=2}^{c-1}\pi^*_{1j}(1-\pi^*_{2j})(\sum_{j'=1}^{j-1}\pi^*_{2j'}) + 2\psi\sum_{j=1}^{c-2}\pi^*_{1j}\pi^*_{2j}\sum_{j'=j+1}^{c-1}(1-\pi^*_{2j'}) \\
&+ \psi^2\sum_{j=1}^{c-1}(\pi^*_{2j})^2(1-\pi^*_{1j}) + 2\psi^2\sum_{j=2}^{c-1}(1-\pi^*_{1j})\pi^*_{2j}\sum_{j'=1}^{j-1}\pi^*_{2j'}\Bigg] - \frac{1}{\psi}\Bigg[\sum_{j=1}^{c-1}\pi^*_{2j}(1-\pi^*_{2j})\Bigg] \\
&= \frac{1}{\psi^2}\Bigg[\psi\sum_{j=1}^{c-1}(1-\pi^*_{2j})\pi^*_{2j}(1-\pi^*_{1j}) + 2\psi\sum_{j=1}^{c-2}\pi^*_{2j}(1-\pi^*_{1j})\sum_{j'=j+1}^{c-1}(1-\pi^*_{2j'}) \\
&- 2\psi\sum_{j=2}^{c-1}\pi^*_{1j}(1-\pi^*_{2j})(\sum_{j'=1}^{j-1}\pi^*_{2j'}) + 2\psi\sum_{j=1}^{c-2}\pi^*_{1j}\pi^*_{2j}\sum_{j'=j+1}^{c-1}(1-\pi^*_{2j'}) \\
&+ \psi\sum_{j=1}^{c-1}\pi^*_{2j}\pi^*_{1j}(1-\pi^*_{2j}) + 2\psi\sum_{j=2}^{c-1}\pi^*_{1j}(1-\pi^*_{2j})(\sum_{j'=1}^{j-1}\pi^*_{2j'})\Bigg] - \frac{1}{\psi}\Bigg[\sum_{j=1}^{c-1}\pi^*_{2j}(1-\pi^*_{2j})\Bigg] \\
&= \frac{1}{\psi}\Bigg[\sum_{j=1}^{c-1}\pi^*_{2j}(1-\pi^*_{2j}) + 2\sum_{j=1}^{c-2}\pi^*_{2j}\sum_{j'=j+1}^{c-1}(1-\pi^*_{2j'}) - \sum_{j=1}^{c-1}\pi^*_{2j}(1-\pi^*_{2j})\Bigg] \\
&= \frac{2}{\psi}\sum_{j=1}^{c-2}\pi^*_{2j}\sum_{j'=j+1}^{c-1}(1-\pi^*_{2j'})\,.
\end{aligned}
$$

Similarly, we can prove

$$K_2 = \frac{2}{\psi}\sum_{j=1}^{c-2}\pi^*_{1j}\sum_{j'=j+1}^{c-1}(1-\pi^*_{1j'})$$

by interchanging the rows and columns. Therefore, we conclude

$$V_{\mathrm{R}} = 2rl(1-l)^2\psi\sum_{j=1}^{c-2}\pi^*_{2j}\sum_{j'=j+1}^{c-1}(1-\pi^*_{2j'}) + 2rl^2(1-l)\psi\sum_{j=1}^{c-2}\pi^*_{1j}\sum_{j'=j+1}^{c-1}(1-\pi^*_{1j'}).$$

Then, we complete the proof of $\psi^2 t_k^2/\omega_k = \lim_{N_+\to\infty}[(\sum_j D_{jk} + 2\sum_{j<j'} D_{jj'k})/N_+]$. □

APPENDIX B
CALCULATIONS FOR SECTION 2.6

For the $2 \times c \times K$ contingency tables, the proposed common cumulative odds ratio estimator and its variance estimator have simpler forms when each stratum consists of a single matched pair. In this appendix, we will simplify the estimators. First, we recall the proposed estimators as follows:

$$\hat{\psi} = \frac{\sum_{k=1}^{K}\sum_{j=1}^{c-1} X^*_{1jk}(n_{2k} - X^*_{2jk})/N_k}{\sum_{k=1}^{K}\sum_{j=1}^{c-1} X^*_{2jk}(n_{1k} - X^*_{1jk})/N_k},$$

$$\hat{\mathrm{Var}}(\hat{\psi}) = \frac{\sum_{k=1}^{K}[\sum_{j=1}^{c-1}\hat{\phi}_{jk}(\hat{\psi}) + 2\sum_{j<j'}^{c-1}\hat{\phi}_{jj'k}(\hat{\psi})]}{[\sum_{k=1}^{K}\sum_{j=1}^{c-1} X^*_{2jk}(n_{1k} - X^*_{1jk})/N_k]^2},$$

where

$$\hat{\phi}_{jk}(\psi) = \frac{n_{1k}n_{2k}}{N_k^2}\{\frac{\psi(n_{1k} - X^*_{1jk})X^*_{2jk}}{n_{1k}}[1 + (\psi - 1)\frac{X^*_{2jk}}{n_{2k}}]$$
$$+\frac{X^*_{1jk}(n_{2k} - X^*_{2jk})}{n_{2k}}[\psi - (\psi - 1)\frac{X^*_{1jk}}{n_{1k}}]\},$$

$$\hat{\phi}_{jj'k}(\psi) = \frac{n_{1k}n_{2k}}{N_k^2}\{\frac{\psi(n_{1k} - X^*_{1j'k})X^*_{2jk}}{n_{1k}}[1 + (\psi - 1)\frac{X^*_{2j'k}}{n_{2k}}]$$
$$+\frac{X^*_{1jk}(n_{2k} - X^*_{2j'k})}{n_{2k}}[\psi - (\psi - 1)\frac{X^*_{1j'k}}{n_{1k}}]\}.$$

Similar to the example discussed in Section 2.6, when each stratum consists of a single matched pair, we can re-express $2 \times c \times K$ contingency tables as $c \times c$ square tables. Let x_{ab} denote the count for the ath row and bth column in a $c \times c$ square table. Then, in a $2 \times c \times K$ contingency table, we have x_{ab} strata with a pattern in which

the response for first level of the explanatory variable is a and the response for the second level of the explanatory variable is b. Also, under the matched pair situation, the term X^*_{ijk} is either 0 or 1 and $n_{ik} = 1$ for all $i \in \{1,2\}$, $j \in \{1,\ldots,c-1\}$, and $k \in \{1,\ldots,K\}$. Under the notation x_{ab}, the $\hat{\psi}$ equals

$$\begin{aligned}\hat{\psi} &= \frac{\sum_{k=1}^{K}\sum_{j=1}^{c-1} X^*_{1jk}(n_{2k} - X^*_{2jk})/N_k}{\sum_{k=1}^{K}\sum_{j=1}^{c-1} X^*_{2jk}(n_{1k} - X^*_{1jk})/N_k} \\ &= \frac{\sum_{k=1}^{K}\sum_{a<b}^{c}(b-a)X_{1ak}X_{2bk}}{\sum_{k=1}^{K}\sum_{a>b}^{c}(a-b)X_{1ak}X_{2bk}} = \frac{\sum_{a<b}(b-a)x_{ab}}{\sum_{a>b}(a-b)x_{ab}}.\end{aligned}$$

For the variance estimator, we first simplify the term $\sum_k \sum_j \hat{\phi}_{jk}(\hat{\psi})$ as

$$\begin{aligned}&\sum_k \sum_j \hat{\phi}_{jk}(\hat{\psi}) \\ =\ & \frac{1}{4}\Bigg[\sum_k \sum_j \hat{\psi}(n_{1k} - X^*_{1jk})X^*_{2jk}[1 + (\hat{\psi} - 1)X^*_{2jk}] \\ &+ \sum_k \sum_j X^*_{1jk}(n_{2k} - X^*_{2jk})[\hat{\psi} - (\hat{\psi} - 1)X^*_{1jk}]\Bigg] \\ =\ & \frac{1}{4}\Bigg[\sum_k \sum_j \hat{\psi}(n_{1k} - X^*_{1jk})X^*_{2jk} + \hat{\psi}(\hat{\psi} - 1)(n_{1k} - X^*_{1jk})X^*_{2jk} \\ &+ \sum_k \sum_j \hat{\psi} X^*_{1jk}(n_{2k} - X^*_{2jk}) - (\hat{\psi} - 1)X^*_{1jk}(n_{2k} - X^*_{2jk})\Bigg] \\ =\ & \frac{1}{4}\Bigg[\hat{\psi}^2 \sum_k \sum_j (n_{1k} - X^*_{1jk})X^*_{2jk} + \sum_k \sum_j X^*_{1jk}(n_{2k} - X^*_{2jk})\Bigg] \\ =\ & \frac{1}{4}\Bigg[\hat{\psi}^2 \sum_{a>b}(a-b)x_{ab} + \sum_{a<b}(b-a)x_{ab}\Bigg].\end{aligned}$$

Also, the term $2\sum_k \sum_{j<j'} \hat{\phi}_{jj'k}$ can simplify as

$$
\begin{aligned}
& 2\sum_{k}\sum_{j<j'}^{c-1}\hat{\phi}_{jj'k} \\
= \; & \frac{2}{4}\Bigg[\hat{\psi}\sum_{k}\sum_{j<j'}[(n_{1k}-X^{*}_{1j'k})X^{*}_{2jk}+\hat{\psi}(\hat{\psi}-1)(n_{1k}-X^{*}_{1j'k})X^{*}_{2jk}X^{*}_{2j'k}] \\
& +\hat{\psi}\sum_{k}\sum_{j<j'}[X^{*}_{1jk}(n_{2k}-X^{*}_{2j'k})-(\hat{\psi}-1)X^{*}_{1jk}(n_{2k}-X^{*}_{2j'k})X^{*}_{1j'k}]\Bigg] \\
= \; & \frac{2}{4}\Bigg[\hat{\psi}\sum_{a>b}\sum_{j=1}^{a-b}(a-b-j)x_{ab}+\hat{\psi}(\hat{\psi}-1)\sum_{a>b}\sum_{j=1}^{a-b}(a-b-j)x_{ab} \\
& +\hat{\psi}\sum_{a<b}\sum_{j=1}^{b-a}(b-a-j)x_{ab}-(\hat{\psi}-1)\sum_{a<b}\sum_{j=1}^{b-a}(b-a-j)x_{ab}\Bigg] \\
= \; & \frac{2}{4}\Bigg[\hat{\psi}^2\sum_{a>b}\sum_{j=1}^{a-b}(a-b-j)x_{ab}+\sum_{a<b}\sum_{j=1}^{b-a}(b-a-j)x_{ab}\Bigg].
\end{aligned}
$$

Then, the $\hat{\mathrm{V}}\mathrm{ar}(\log\hat{\psi})$ is

$$
\begin{aligned}
&\hat{\mathrm{V}}\mathrm{ar}(\log\hat{\psi}) \\
&= \left[\frac{\hat{\psi}^2\sum_{a>b}(a-b)x_{ab}+\sum_{a<b}(b-a)x_{ab}+2\hat{\psi}^2\sum_{a>b}\sum_{j=1}^{a-b}(a-b-j)x_{ab}}{[\sum_{a>b}(a-b)x_{ab}]^2}\right. \\
&\quad \left.+\frac{2\sum_{a<b}\sum_{j=1}^{b-a}(b-a-j)x_{ab}}{[\sum_{a>b}(a-b)x_{ab}]^2}\right]\times\frac{[\sum_{a>b}(a-b)x_{ab}]^2}{[\sum_{a<b}(b-a)x_{ab}]^2} \\
&= \frac{\hat{\psi}^2[\sum_{a>b}(a-b)x_{ab}+2\sum_{a>b}\sum_{j=1}^{a-b}(a-b-j)x_{ab}]}{[\sum_{a<b}(b-a)x_{ab}]^2} \\
&\quad +\frac{\sum_{a<b}(b-a)x_{ab}+2\sum_{a<b}\sum_{j=1}^{b-a}(b-a-j)x_{ab}}{[\sum_{a<b}(b-a)x_{ab}]^2} \\
&= \frac{\sum_{a>b}(a-b)x_{ab}+2\sum_{a>b}\sum_{j=1}^{a-b}(a-b-j)x_{ab}}{[\sum_{a>b}(a-b)x_{ab}]^2} \\
&\quad +\frac{\sum_{a<b}(b-a)x_{ab}+2\sum_{a<b}\sum_{j=1}^{b-a}(b-a-j)x_{ab}}{[\sum_{a<b}(b-a)x_{ab}]^2} \\
&= \frac{\sum_{a>b}(a-b)^2x_{ab}}{[\sum_{a>b}(a-b)x_{ab}]^2}+\frac{\sum_{a<b}(b-a)^2x_{ab}}{[\sum_{a<b}(b-a)x_{ab}]^2}.
\end{aligned}
$$

The simpler formula equals the Agresti and Lang (1993) proposed estimator and variance estimator for a proportional odds model for subject–specific distributions.

REFERENCES

Abramowitz, M., & Stegun, I. (1972). *Handbook of Mathematical Functions.* Dover, New York.

Agresti, A. (1984). *Analysis of Ordinal Categorical Data.* Wiley, New York.

Agresti, A. (1990). *Categorical Data Analysis.* Wiley, New York.

Agresti, A., & Lang, J. B. (1993). A proportional odds model with subject-specific effects for repeated ordered categorical responses. *Biometrika, 80,* 527–534.

Andersen, E. B. (1973). *Conditional Inference and Models for Measuring.* Mentalhygienisk Forlag, Copenhagen.

Andersen, E. B. (1980). *Discrete Statistical Methods with Social Science Applications.* North-Holland, New York.

Barndorff-Nielsen, O. E., & Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics.* Chapman and Hall, London.

Begg, C. B., & Gray, R. (1984). Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika, 71,* 11–18.

Birch, M. W. (1965). The detection of partial association II: the general case. *Journal of the Royal Statistical Society, Series B, 27,* 111–124.

Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* MIT Press, Cambridge, Massachusetts.

Breslow, N. E. (1981). Odds ratio estimators when the data are sparse. *Biometrika, 68,* 74–84.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association, 88,* 9–25.

Breslow, N. E., & Day, N. E. (1980, 1987). *Statistical Methods in Cancer Research, Vol. I: The Analysis of Case-Control Studies; Vol. II: The Design and Analysis of Cohort Studies.* IARC, Lyon.

Cochran, W. G. (1954). Some methods of strengthening the common χ^2 tests. *Biometrics, 10,* 417–451.

Dale, J. R. (1986). Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics, 42,* 909–917.

De Bruijn, N. G. (1961). *Asymptotic Methods in Analysis.* Dover, New York.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions* (second edition). Wiley, New York.

Gart, J. J. (1962). On the combination of the relative risks. *Biometrics, 18*, 594–600.

Greenland, S. (1989). Generalized Mantel–Haenszel estimators for K $2 \times J$ tables. *Biometrics, 45*, 183–191.

Guilbaud, O. (1983). On the large–sample distribution of the Mantel–Haenszel odds–ratio estimator. *Biometrics, 39*, 523–525.

Hauck, W. W. (1979). The large sample variance of the Mantel–Haenszel estimator of a common odds ratio. *Biometrics, 35*, 817–819.

Koch, G. G., Imrey, P. B., Singer, J. M., Atkinson, S. S., & Stokes, M. E. (1985). *Analysis of Categorical Data.* Presses de L'Univ. Montreal, Montreal.

Kuritz, S. J., Landis, J. R., & Koch, G. G. (1988). A general overview of Mantel–Haenszel methods: applications and recent developments. *Annual Review of Public Health, 9*, 123–160.

Landis, J. R., Heyman, E. R., & Koch, G. G. (1988). Average partial association in three–way contingency tables: A review and discussion of alternative tests. *International Statistical Review, 46*, 237–254.

Lang, J. B. (1992). On model fitting for multivariate polytomous response data. Dissertation, University of Florida.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks.* Holden–Day, San Francisco.

Lehmann, E. L. (1983). *Theory of Point Estimation.* Wiley, New York.

Liang, K. Y. (1985). Odds ratio inference with dependent data. *Biometrika, 72*, 678–682.

Liang, K. Y. (1987). Extended Mantel–Haenszel estimating procedure for multivariate logistic regression models. *Biometrics, 43*, 289–299.

Liang, K. Y., & Self, S. G. (1985). Tests for homogeneity of odds ratio when the data are sparse. *Biometrika, 72*, 353–358.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 13–22.

Liu, Q., & Pierce, D. A. (1993). Heterogeneity in Mantel–Haenszel–type models. *Biometrika, 80*, 543–556.

Liu, Q., & Pierce, D. A. (1994). A Note on Gauss–Hermite Quadrature. *Biometrika, 81*, 624–629.

Madsen, M. (1976). Statistical analysis of multiple contingency tables. Two examples. *Scandinavian Journal of Statistics, 3*, 97–106.

Mantel, N. (1963). Chi-square tests with one degree of freedom: extensions of the Mantel–Haenszel procedure. *Journal of the American Statistical Association, 58*, 690–700.

Mantel, N., & Byar, D. P. (1978). Marginal homogeneity, symmetry and independence. *Communications in Statistics: Theory and Methods, A7*, 953–976.

Mantel, N., & Fleiss, J. L. (1980). Minimum expected cell size requirements for the Mantel–Haenszel one degree of freedom chi-square test and a related rapid procedure. *American Journal of Epidemiology, 112*, 129–134.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719–748.

McCullagh, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika, 65*, 413–418.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B, 42*, 109–142.

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models* (second edition). Chapman and Hall, New York.

Mickey, R. M., & Elashoff, R. M. (1985). A Generalization of the Mantel–Haenszel estimator of partial association for $2 \times J \times K$ tables. *Biometrics, 41*, 623–635.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1973). *Introduction to the Theory of Statistics* (third edition). McGraw–Hill, New York.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrika, 16*, 1–22.

Noether, G. E. (1955). Use of the range instead of the standard deviation. *Journal of the American Statistical Association, 50*, 1040–1055.

Plackett, R. L. (1981). *The Analysis of Categorical Data* (second edition). Charles Griffin House, London.

Puri, M. L., & Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis.* Wiley, New York.

Randles, R. H., & Wolfe, D. A. (1991). *Introduction to The Theory of Nonparametric Statistics.* Krieger Drive, Florida.

Robins, J. N., Breslow, N. E., & Greenland, S. (1986). Estimators of the Mantel–Haenszel variance consistent in both sparse data and large-strata limiting models. *Biometrics, 42*, 311–323.

Sato, T. (1991). An estimating equation approach for the analysis of case-control studies with exposure measured at several levels. *Statistics in Medicine, 10*, 1037–1042.

Simon, G. A. (1978). Efficacies of measures of association for ordinal contingency tables. *Journal of the American Statistical Association*, *73*, 545–551.

Stuart, A. (1953). The estimation and comparison of strengths of association in contingency tables. *Biometrika*, *40*, 105–110.

Tarone, R. E. (1985). On heterogeneity tests based on efficient scores. *Biometrika*, *72*, 91–95.

Tierney, L., & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, *81*, 82–86.

Woolf, B. (1955). On estimating the relation between blood group and disease. *Annals of Human Genetics*, *19*, 251–253.

Yanagawa, T., & Fujii, Y. (1990). Homogeneity test with a generalized Mantel–Haenszel estimator for L $2 \times K$ contingency tables. *Journal of the American Statistical Association*, *85*, 744–748.

Yanagawa, T., & Fujii, Y. (1994). Projection method Mantel–Haenszel estimator for K $2 \times J$ tables. *Journal of the American Statistical Association*, in press.

BIOGRAPHICAL SKETCH

I-Ming Liu was born on December 20, 1964 in Taipei, Taiwan. After graduating from high school in 1983, she joined National Chengchi University, Taipei. She received a Bachelor of Business Administration degree in June, 1987, from the Department of Statistics. Before she came to the University of Florida to pursue her Ph.D. studies, she obtained her Master of Statistics degree from Iowa State University, Ames, in June, 1989. She expects to get her Ph.D. in May, 1995.

As a graduate student, she was a research assistant for three years working for Professors Mark Yang, Alan Agresti, and Ramon Littell. During the last year of her graduation, she was a teaching assistant.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Alan G. Agresti, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Myron N. Chang
Associate Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Malay Ghosh
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ramon C. Littell
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James J. Algina
Professor of Foundations of Education

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May 1995

Dean, Graduate School